NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Space Programs Summary No. 37-32, Volume IV

for the period February 1, 1965 to March 31, 1965

Supporting Research and Advanced Development

jpl
JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

April 30, 1965

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Space Programs Summary No. 37-32, Volume IV

for the period February 1, 1965 to March 31, 1965

Supporting Research and Advanced Development

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

April 30, 1965

Preface

The *Space Programs Summary* is a six volume, bimonthly publication designed to report on JPL space exploration programs, and related supporting research and advanced development projects. The subtitles of all volumes of the *Space Programs Summary* are:

- Vol. I. The Lunar Program (Confidential)
- Vol. II. The Planetary–Interplanetary Program (Confidential)
- Vol. III. The Deep Space Network (Unclassified)
- Vol. IV. Supporting Research and Advanced Development (Unclassified)
- Vol. V. Supporting Research and Advanced Development (Confidential)
- Vol. VI. Space Exploration Programs and Space Sciences (Unclassified)

The *Space Programs Summary*, Vol. VI consists of an unclassified digest of appropriate material from Volumes I, II, and III; an original presentation of the JPL (1) quality assurance and reliability efforts, and (2) environmental- and dynamic-testing facility-development activities; and a reprint of the space science instrumentation studies of Volumes I and II.

W. H. Pickering

W. H. Pickering, *Director*
Jet Propulsion Laboratory

Space Programs Summary No. 37-32, Volume IV

Prepared under Contract No. NAS 7-100, National Aeronautics & Space Administration

Contents

Contents (Cont'd)

Contents (Cont'd)

Contents (Cont'd)

Contents (Cont'd)

TELECOMMUNICATIONS DIVISION

Contents (Cont'd)

SYSTEMS DIVISION

I. Systems Analysis

A. Effects of the Moon's Gravitational Harmonics on a Lunar Orbiter

F. M. Sturms, Jr.

This article gives some preliminary results of a study to examine the sensitivity of orbit elements and tracking station range rate to values of the harmonic coefficients in the lunar gravitational field. It is hoped that these results will give insight into which coefficients can be determined by tracking a lunar orbiter and what methods of orbit determination are applicable to a selenodesy effort for a lunar orbiter.

To implement the study, a special version of the JPL space trajectories program was developed, with the lunar gravitational potential in terms of spherical harmonics in the form:

$$U = \frac{\mu}{r} - \frac{\mu}{r}\sum_{n=1}^{N} J_n \left(\frac{R}{r}\right)^n P_n(\sin\phi) + \frac{\mu}{r}\sum_{n=1}^{N}\sum_{m=1}^{n}\left(\frac{R}{r}\right)^n P_{n,m}(\sin\phi)\,(C_{n,m}\cos m\lambda + S_{n,m}\sin m\lambda).$$

A subroutine for obtaining the Cartesian perturbation components for $N \leq 10$ was coded; it replaces the standard subroutine in terms of the moments of inertia (A, B, C) about the principal axes. The program was checked out by comparing runs with values of R, J_2, and $C_{2,2}$ equivalent to values of A, B, and C on the standard trajectory program.

1. Nominal Values of Coefficients

The nominal values of J_1, $C_{1,1}$, $C_{2,1}$, $S_{1,1}$, $S_{2,1}$, and $S_{2,2}$ are zero, from the definition of the Moon-centered coordinate system. Non-zero values of J_1, $C_{1,1}$, and $S_{1,1}$ indicate a coordinate system that is not centered at the Moon's center of mass; non-zero values of $C_{2,1}$ and $S_{2,1}$ indicate coordinate axes that are not along the principal axes of inertia; and a non-zero value of $S_{2,2}$ indicates longitude not measured from the principal x-axis. Since it is not clear at this time whether they should be determined directly or by the equivalent determination of the Moon's ephemeris or coordinate system, these coefficients are not examined in this study.

Values of J_2 and $C_{2,2}$ are fairly well-known. The nominal values used in this study are equivalent to the values of A, B, and C in Ref. 1: $J_2 = 207.11 \times 10^{-6}$ and

$C_{2,2} = 20.716 \times 10^{-6}$. Values of J_2 and $C_{2,2}$ are increased by 10^{-6} for perturbed runs.

The value of the important coefficient J_3 is not well-known; the nominal value is taken as zero, as are the nominal values of all other coefficients except J_2 and $C_{2,2}$. The perturbed values of J_3 are taken as $\pm 10^{-4}$. Runs have not yet been obtained for the perturbed values of other coefficients.

2. Nominal Orbit

The selenographic initial conditions for the nominal orbit (apoapsis injection) were:

$$t = 07{:}00 \text{ GMT, 8-6-66}$$

$$r = 2663 \text{ km}$$

$$\phi = 0 \text{ deg}$$

$$\theta = 160.7 \text{ deg}$$

$$V = 1.2154 \text{ km/sec}$$

$$\Gamma = 0 \text{ deg}$$

$$\Sigma = 111.3 \text{ deg}$$

with $\mu = 4902.7779$ km^3/sec^2. The resulting orbital conditions were:

$$a_0 = 2239.9188 \text{ km}$$

$$e_0 = 0.18888214$$

$$\text{periapsis altitude} = 77.8 \text{ km}$$

$$\text{period} = 158.55 \text{ min}$$

$$i_0 = 21.179 \text{ deg}$$

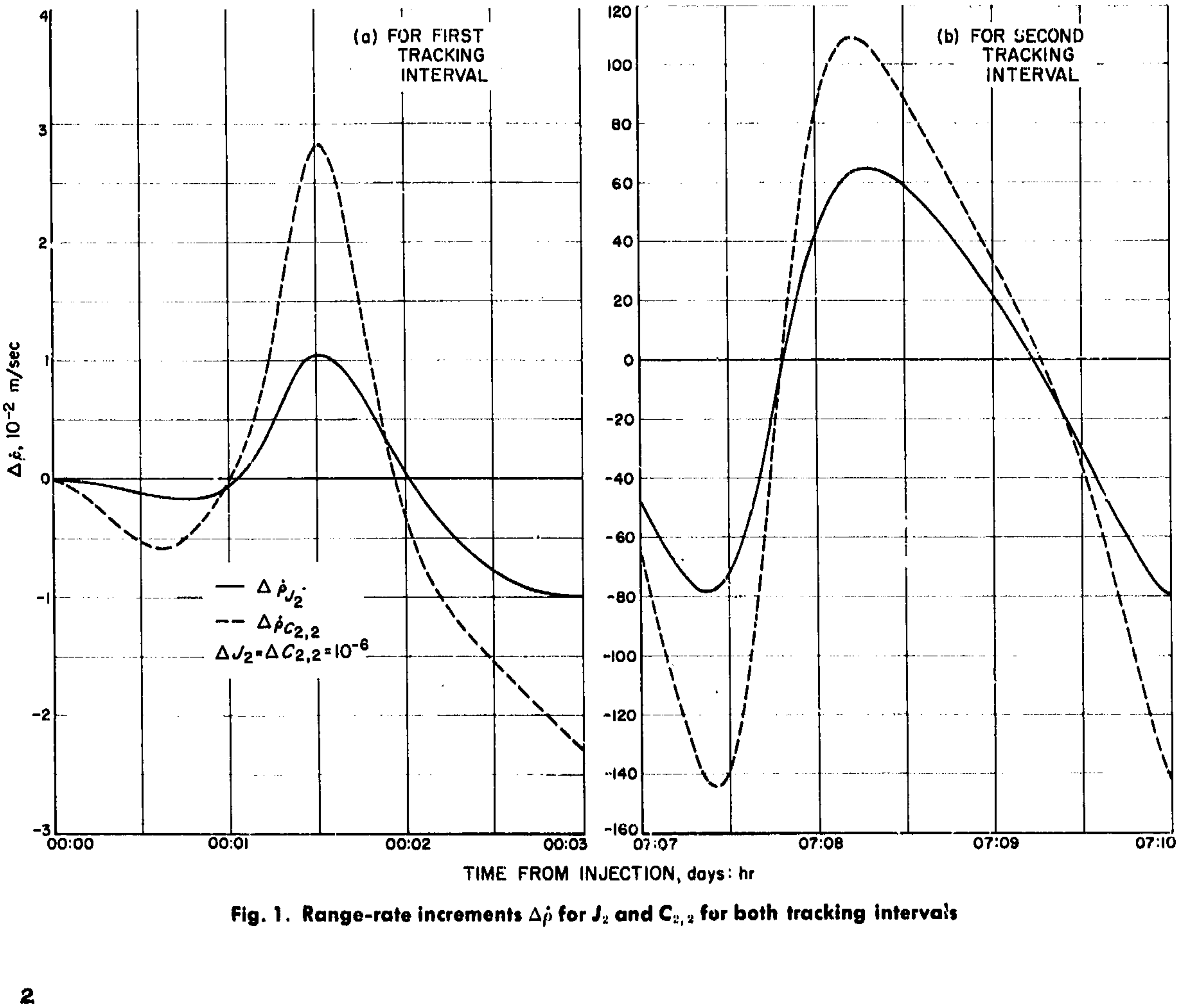


Fig. 1. Range-rate increments $\Delta\dot{\rho}$ for J_2 and $C_{2,2}$ for both tracking intervals

All orbits were obtained by Moon-centered Encke integration, with a 60-sec step size.

3. Effects on Range Rate

The range rate $\dot{\rho}$ was obtained for two 3-hr periods for the Goldstone Tracking Station. The periods are from injection to 3 hr after injection and from 7 days, 7 hr after injection to 7 days, 10 hr after injection. Station printouts were obtained at 2-min intervals.

Fig. 1 shows $\Delta\dot{\rho}$ (perturbed range rate minus nominal range rate) for a ΔJ_2 and $\Delta C_{2,2}$ of 10^{-6} for both tracking intervals. The maximum residuals for the first tracking interval are 0.010 and 0.030 m/sec, respectively. For the second tracking interval, the residuals increase to 0.8 and 1.4 m/sec, respectively. The magnitudes of the range-rate residuals are within the Deep Space Network capability; however, the J_2 and $C_{2,2}$ residuals are almost proportional and may be difficult to separate. In order to describe the separability of the two, Fig. 2 shows the quantity

$$\Delta\dot{\rho}_{J_2} - R\Delta\dot{\rho}_{C_{2,2}},$$

where the proportionality constant R was chosen to equalize the positive peaks in Fig. 1. The differences are indicative of the separability of J_2 and $C_{2,2}$ by the orbit-determination process. After 7 days, the effective maximum residual is 0.11 m/sec.

Fig. 3 shows the effects of both positive and negative values of ΔJ_3 on $\Delta\dot{\rho}$ in the two tracking intervals. The value of ΔJ_3 is $\pm 10^{-4}$. The effects of positive and negative values are nearly opposite, indicating linearity in the range of values. Kaula[1] has shown that the magnitude of J_3 may be as high as 93×10^{-6}, based on equal stress assumptions for the Earth and Moon. As shown in Fig. 3,

[1]Kaula, W. M., *Calculation of Perturbations of Lunar Orbiters*, U.S. Govt. Memorandum, December 13, 1963 (Unpublished).

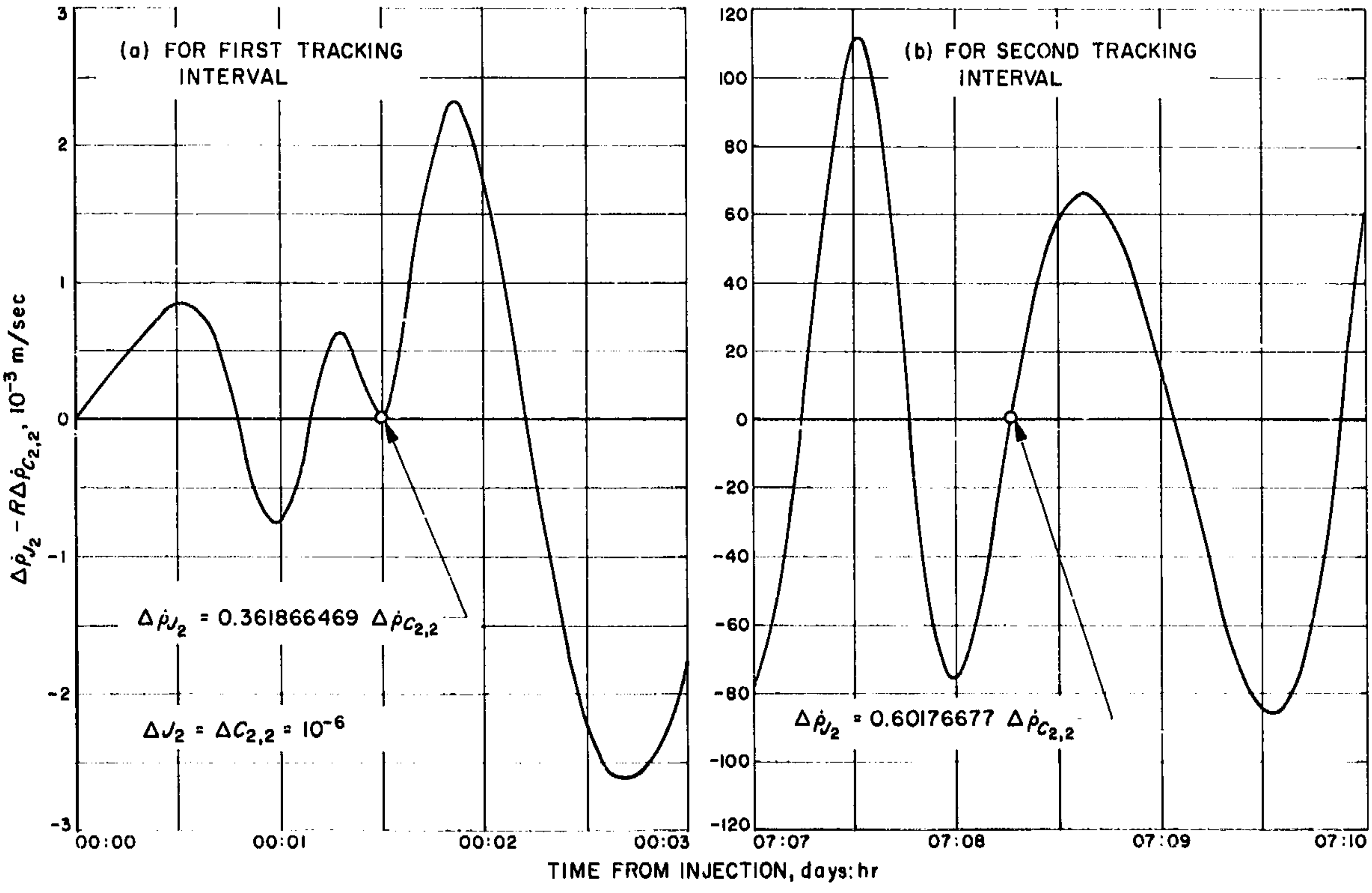


Fig. 2. Differences in J_2 and $C_{2,2}$ range-rate increments $\Delta\dot{\rho}$ for both tracking intervals, using proportionality constant R

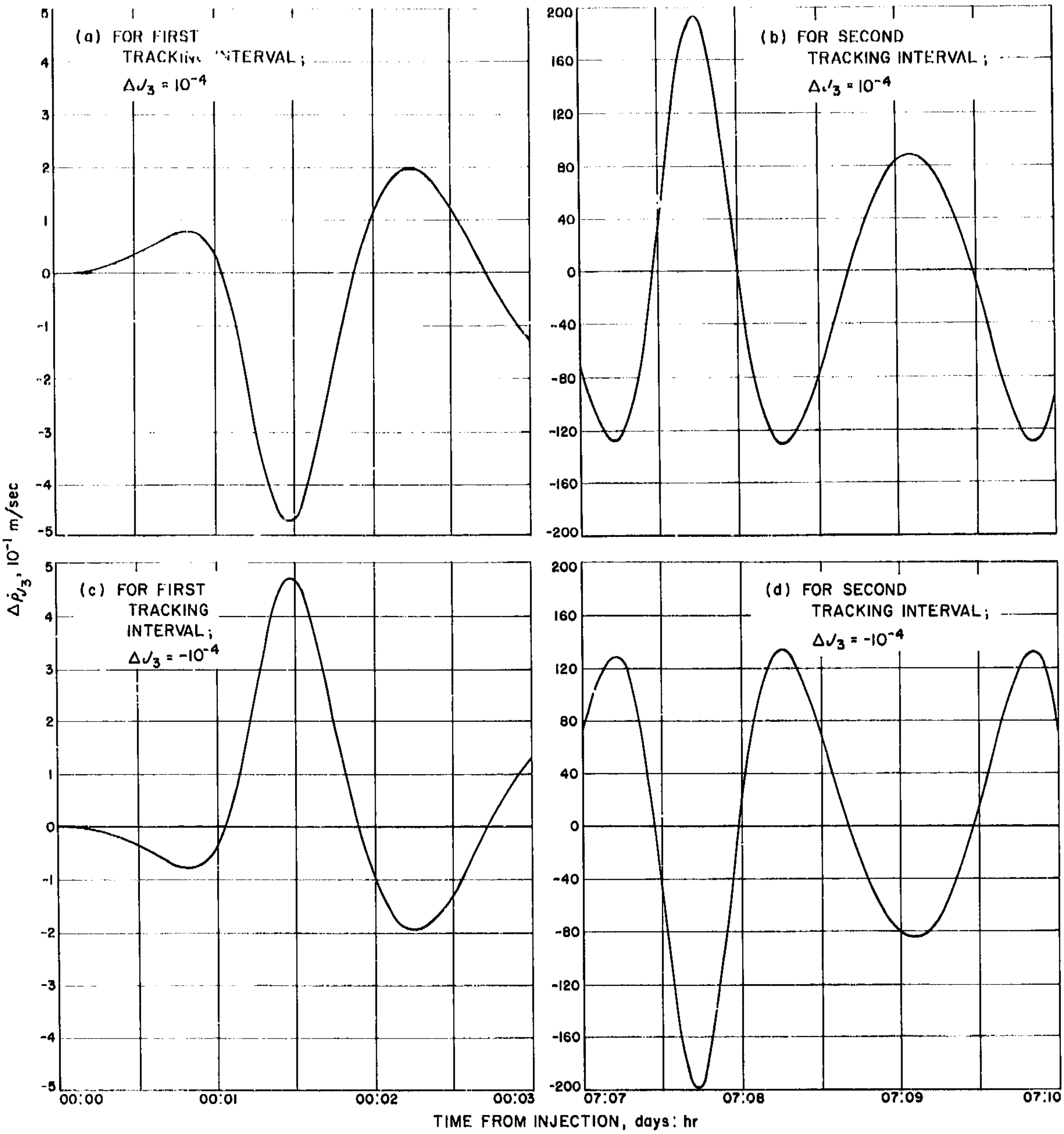


Fig. 3. Range-rate increments $\Delta\dot{\rho}$ for J_3 for both tracking intervals

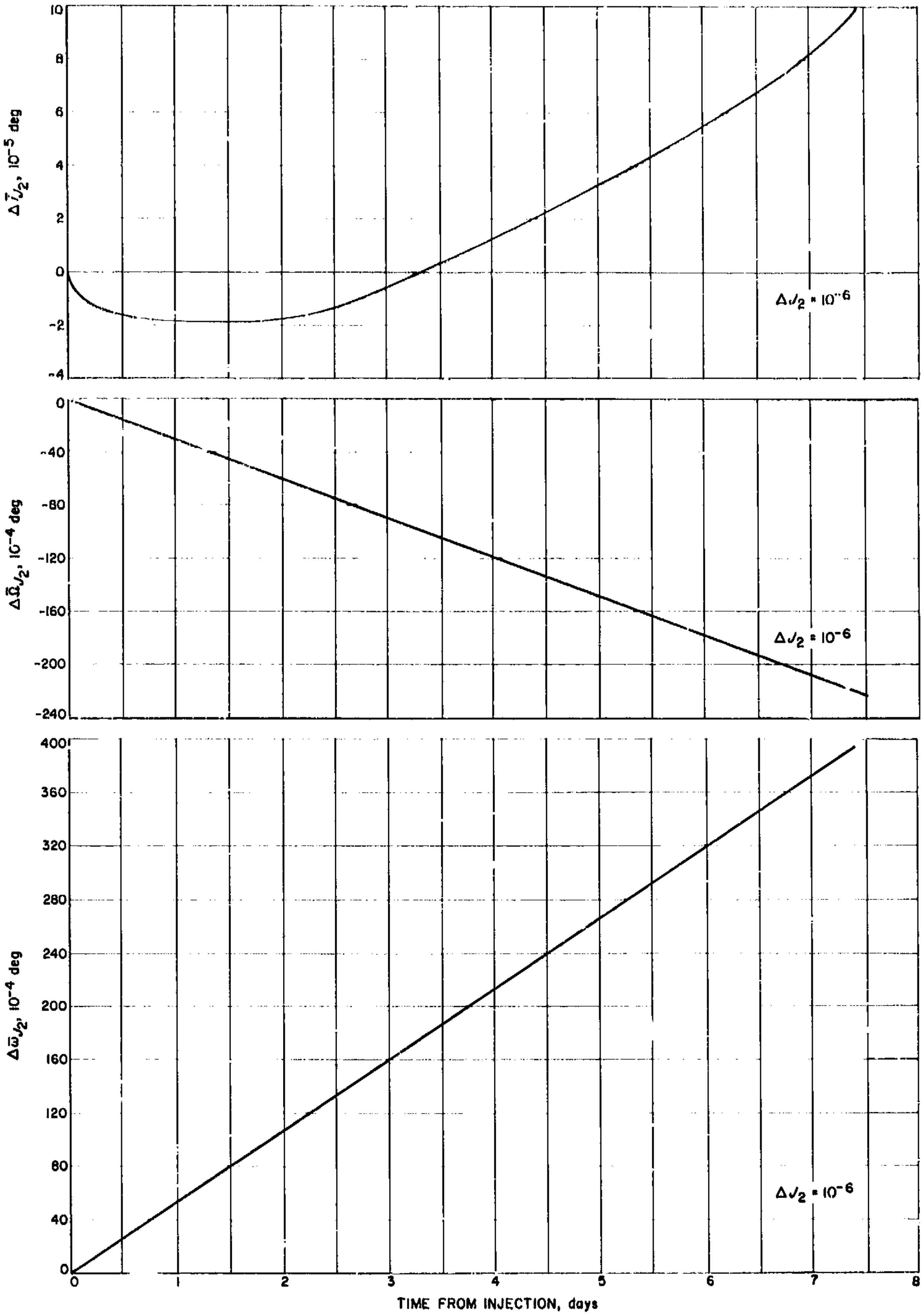


Fig. 4. Increments of mean inclination $\bar{i}$, longitude of node $\bar{\Omega}$, and argument of periapsis $\bar{\omega}$ for J_2 for the first 7 days after injection

Table 1. Values of increments in several osculating orbit elements for $\Delta J_2 = 10^{-6}$

Time from injection, day:hr:min	Δe_{J_2}, $\times 10^{-7}$	Δi_{J_2}, 10^{-6} deg	$\Delta\Omega_{J_2}$, 10^{-4} deg	$\Delta\omega_{J_2}$, 10^{-4} deg	$(\Delta r_p)_{J_2}$, 10^{-3} km
First tracking interval					
00:00:10	−0.3	0.1	0	−0.5	0.1
20	2.2	−0.5	0	−0.9	−0.4
30	2.8	−0.6	−0.2	−1.1	−0.6
40	5.0	−1.3	−0.4	−1.2	−1.2
50	8.1	−1.9	−0.8	−0.8	−1.6
00:01:00	10.4	−1.3	−1.2	−0.2	−2.0
10	14.9	0.2	−1.6	0.9	−2.3
20	18.5	0.2	−1.7	3.1	−2.8
30	15.3	−0.4	−1.7	5.2	−2.4
40	10.7	−1.4	−2.1	6.2	−1.9
50	6.8	−1.9	−2.5	6.6	−1.5
00:02:00	5.1	−1.4	−2.9	6.8	−1.1
10	2.6	−1.0	−3.1	7.0	−0.6
20	1.5	−0.4	−3.3	6.7	−0.4
30	−0.6	0	−3.3	6.3	0.1
40	0	0	−3.4	5.8	0.1
50	0.6	−0.3	−3.3	5.2	−0.2
00:03:00	0.8	−0.7	−3.3	4.9	−0.1

Time from injection, day:hr:min	Δe_{J_2}, $\times 10^{-7}$	Δi_{J_2}, 10^{-6} deg	$\Delta\Omega_{J_2}$, 10^{-4} deg	$\Delta\omega_{J_2}$, 10^{-4} deg	$(\Delta r_p)_{J_2}$, 10^{-3} km
Second tracking interval					
07:07:00	−3.0	9.0	−216.2	387.718	0.8
10	0	9.2	−216.6	387.664	0.2
20	4.0	9.7	−216.8	388.245	−0.4
30	7.6	10.7	−217.1	389.288	−0.9
40	10.9	10.8	−217.0	391.08	−1.3
50	8.8	9.8	−217.1	393.112	−1.0
07:08:00	4.8	8.8	−217.6	394.536	−0.5
10	1.1	8.7	−218.2	395.105	−0.1
20	−2.2	9.7	−218.6	395.400	0.4
30	−3.8	10.1	−218.6	395.313	0.9
40	−5.0	10.6	−218.7	394.802	1.0
50	−5.2	10.7	−218.6	394.255	1.0
07:09:00	−6.2	10.9	−218.7	393.833	1.3
10	−6.0	10.2	−218.6	393.414	1.3
20	−4.4	9.5	−218.8	393.172	0.8
30	−3.1	9.0	−219.0	392.894	0.6
40	−1.9	9.3	−219.3	393.176	0.5
50	−0.2	9.7	−219.7	393.446	0.2
07:10:00	4.6	9.8	−221.2	394.008	−0.4

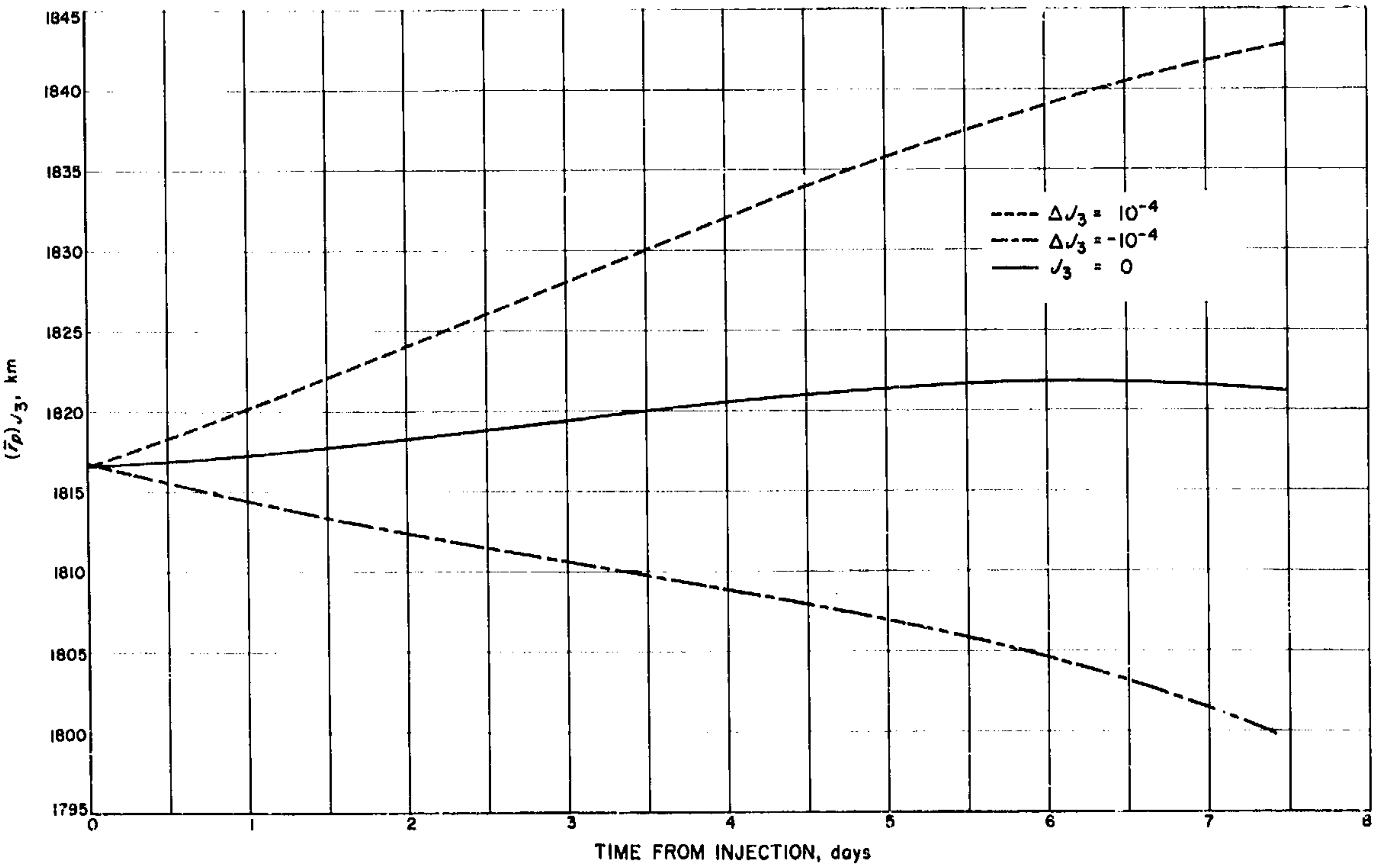


Fig. 5. Mean radius of periapsis $\bar{r}_p$ for J_3 for the first 7 days after injection

this value would give residuals which could be easily detected by the Deep Space Network.

4. Effects on Orbit Elements

Another method for determining the values of the harmonic coefficients is based on the perturbations of the osculating orbit elements or the mean elements. The values of increments in several osculating orbit elements (perturbed values minus nominal values) due to a ΔJ_2 of 10^{-6} are given in Table 1 at 10-min intervals; these elements are as follows: eccentricity e, inclination i, longitude of node Ω, argument of periapsis ω, and radius of periapsis r_p.

It is likely that mean elements will be used in an orbit-determination method rather than osculating elements. Values of the mean orbit elements were computed for the first 7 days after injection. The mean value was obtained by fairing a smooth curve through osculating values at 4-hr intervals. Values of $\Delta\bar{e}_{J_2}$ and $(\Delta\bar{r}_p)_{J_2}$ for an increment $\Delta J_2 = 10^{-6}$ were very erratic and are not given here. Values of $\Delta\bar{i}_{J_2}$, $\Delta\bar{\Omega}_{J_2}$, and $\Delta\bar{\omega}_{J_2}$ are shown in Fig. 4. These plots are similarly smoothly faired through points including short-period terms.

The marked effect of J_3 on $\bar{r}_p$ is shown in Fig. 5, which compares $\bar{r}_p$ values for 7 days for $\Delta J_3 = \pm 10^{-4}$ to values for the nominal $J_3 = 0$. This sensitivity makes $\bar{r}_p$ important in determining J_3.

B. Coordinates in the Theory of Relativity

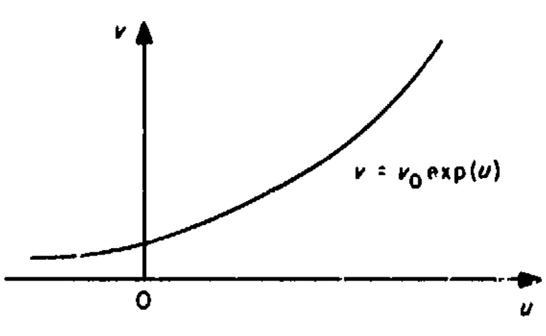
H. Lass

In recent years there have been various criticisms of Einstein's general theory of relativity, most of which concern the use of generalized coordinates and the assumption of the covariance of Einstein's field equation to arbitrary coordinate transformations. Indeed, the Russian physicist Fock (Ref. 2) argues that a preferred coordinate system is required to exhibit certain intrinsic properties of space and time, and he chooses the harmonic conditions $g^{\mu\nu}x^i{}_{,\mu\nu} = 0$, $i = 1, 2, 3, 4$, of de Donder to determine his preferred coordinate system. In any case, however, the precise meaning of these coordinates in terms of physical measurements is ignored by Fock, and Einstein's fundamental principle of the covariance of the laws of physics is destroyed.

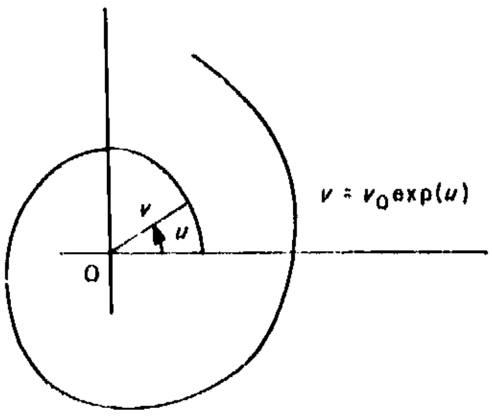
P. Bridgman's critique of Einstein's general theory of relativity is that Einstein failed to carry over to the general theory of relativity the operational methods of defining clocks and lengths which he had so aptly applied to the special theory of relativity. In effect, Bridgman criticizes Einstein for failing to operationally define what is meant by a coordinate system. Einstein's answer to Bridgman's critiques (Ref. 3) was as follows: "In order to be able to consider a logical system as physical theory, it is not necessary to demand that all of its assertions can be independently interpreted and 'tested' operationally, de facto. This has never yet been achieved by any theory and can not at all be achieved. In order to be able to consider a theory as a physical theory, it is only necessary that it implies empirically testable assertions in general."

In this discussion, we wish to show that, under certain conditions, one can define the coordinates of general relativity theory from an operational point of view. We will negate the argument given by various students of relativity that, if the coordinates involved in the Schwarzchild line element for a point mass (or the Sun) yield an advance in the perihelion of Mercury, how can we be sure that a rotation has not in some way been introduced which just accounts for this advance in perihelion?

Now it should be clear that, if one solves a system of differential equations involving certain coordinates, the path defined by the solution depends on the physical or geometrical definitions of these coordinates. For example, consider a law of physics given by $dv/du = v$ whose solution is $v = v_0 \exp(u)$. If we interpret u and v as the abscissa and ordinate, respectively, of Euclidean 2-space, we obtain the following curve:



On the other hand, if this law of physics is interpreted by a two-dimensional observer who can only visualize a polar coordinate system, the path $v = v_0 \exp(u)$ is given by:



Thus, it should be clear that the paths given by the geodesic equations

$$\frac{d^2x^i}{ds^2} + \Gamma^i_{jk}\frac{dx^j}{ds}\frac{dx^k}{ds} = 0 \tag{1}$$

have no physical significance unless the coordinates x^i, $i = 1, 2, 3, 4$, are operationally defined.

Our first example involving the operational definition of coordinates in the theory of relativity will concern itself with a frame of reference rotating at a constant angular speed relative to an inertial frame of reference. Let 0-x-y-z be an inertial frame of reference,

$$\begin{aligned} ds^2 &= c^2dt^2 - (dx^2 + dy^2 + dz^2) \\ &= c^2dt^2 - (dr^2 + r^2d\theta^2 + dz^2). \end{aligned}$$

Now consider the rotating frame of reference centered at 0 and moving with angular speed ω in the sense that any point fixed on the rotating frame has an angular speed ω as observed from the inertial frame of reference. The radius of the rotating frame ("merry-go-round") is assumed finite, $r = r_0$, such that $\omega r_0 < c$. Let us assume that the rotating frame of reference is made of glass so that observers on this frame of reference can look down and note the inertial coordinates x-y-t of an event. We will neglect the z-component and consider only the coordinates r, θ, t of the inertial observers. An event which takes place at r, θ, t is noted by an observer in the rotating frame of reference (noninertial frame), and he designates the "coordinate" time by $T = t$ and the "radial" coordinate by $R = r$. Now the motion of any point fixed in the noninertial frame is given by $\theta = \theta_0 + \omega t$, as observed by the inertial observers. Having chosen a fixed reference direction, the motion of this same point is given by $\phi = \phi_0$. For each θ_0, there exists a ϕ_0 such that $\phi_0 = f(\theta_0)$, and

$$\begin{aligned} \phi &= f(\theta - \omega t), \\ \theta &= \omega t + g(\phi) = \omega T + g(\phi), \\ d\theta &= \omega dT + g'(\phi)\,d\phi. \end{aligned} \tag{2}$$

Hence, the metric for the rotating coordinate system is given by

$$\begin{aligned} ds^2 &= c^2\left(1 - \frac{\omega^2R^2}{c^2}\right)dT^2 - dR^2 - R^2g'^2(\phi)\,d\phi^2 \\ &\quad - 2\omega R^2g'(\phi)\,d\phi\,dT - dZ^2. \end{aligned} \tag{3}$$

From symmetry conditions, the components of the metric tensor $\bar{g}_{ij}$ should be independent of ϕ, so that $g'(\phi) = 1$ in order that $\theta = \phi$ for $\omega = 0$. Hence, we have obtained the well-known transformations

$$\begin{aligned} t &= T \\ r &= R \\ \theta &= \phi + \omega T \\ z &= Z \end{aligned} \tag{4}$$

from an operational point of view, with

$$\begin{aligned} ds^2 &= c^2\left(1 - \frac{\omega^2R^2}{c^2}\right)dT^2 - dR^2 - R^2d\phi^2 \\ &\quad - 2\omega R^2d\phi dT - dZ^2 \end{aligned} \tag{5}$$

for $\omega R < c$. Since proper time is defined by ds/c, we see that, for an observer fixed in the rotating frame of reference, $dR = d\phi = dZ = 0$,

$$\frac{ds}{c} = d\tau = \left(1 - \frac{\omega^2R^2}{c^2}\right)^{1/2} dT, \tag{6}$$

and the rate of "atomic clocks," or proper time, is a function of the radial coordinate of the observer.

It should be noted that, for two points fixed on the rotating frame, the coordinate time for light to go from P_1 to P_2 is given by (setting $ds^2 = 0$):

$$T_2 - T_1 = \int_{T_1}^{T_2} dT$$

$$= \int_{P_1}^{P_2} \frac{\omega R^2 |d\Theta| + \left[\omega^2 R^4 d\Theta^2 + c^2\left(1 - \frac{\omega^2 R^2}{c^2}\right)(dR^2 + R^2 d\Theta^2 + dZ^2)\right]^{1/2}}{c^2\left(1 - \frac{\omega^2 R^2}{c^2}\right)}. \tag{7}$$

If, at T_1, P_1 sends a light signal to P_2, it will be received at P_2 at coordinate time T_2. From Eq. (7), it follows that, if another light signal is sent from P_1 at coordinate time $T_1 + dT$, it will be received at P_2 at coordinate time $T_2 + dT$, so that coordinate time is universal; i.e., all observers agree on the meaning of dT.

Now, suppose we are interested in the distance between two points $P_1(R_1, \Theta_1, Z = 0)$ and $P_2(R_2, \Theta_1, Z = 0)$. One usually sets $d\Theta_1 = 0$ and $dT = 0$ and finds that

$$-ds^2 = dR^2,$$

so that

$$L = \int_{R_1}^{R_2} |dR| = |R_2 - R_1|.$$

In the special theory of relativity, Einstein has shown that the distance between two points involves clocks as well as meter sticks (simultaneity of two events). Since the rate of atomic clocks (true time) depends on the position of the observer, setting $dT = 0$ is not equivalent to making simultaneous marks at P_1 and P_2. However, by means of light signals one can speak of the distance from P_1 to P_2, say $L(P_1, P_2)$. We will find that, in general, $L(P_1, P_2) \neq L(P_2, P_1)$. Let $P_1(R_1, \Theta_1)$ send a light signal to $P_2(R_2, \Theta_2)$, and let the light signal be reflected back to P_1. The coordinate time for the journey of the light signal is seen to be (Eq. 7):

$$T = 2(T_2 - T_1).$$

The proper time is given by:

$$\tau = 2(T_2 - T_1)\left(1 - \frac{\omega^2 R_1^2}{c^2}\right)^{1/2},$$

so that P_1 defines the distance to P_2 by

$$L(P_1, P_2) = c(T_2 - T_1)\left(1 - \frac{\omega^2 R_1^2}{c^2}\right)^{1/2}, \tag{8}$$

and similarly

$$L(P_2, P_1) = c(T_2 - T_1)\left(1 - \frac{\omega^2 R_2^2}{c^2}\right)^{1/2}.$$

As a second example involving the operational point of view of defining coordinates, we will find the metric due to a "uniform" gravitational field. Let us consider a massive (infinite) body at $X = \infty$ which produces a one-dimensional gravitational field in the X-direction (a strictly idealized condition). By means of a rocket, we assume that a frame of reference 0-X-Y-Z exists and is at rest relative to the massive body. At the origin 0 of this coordinate system, light signals are sent to the fixed observers of the 0-X-Y-Z frame of reference at unit "time" intervals. The fixed observers set up "coordinate" clocks to run at the same rate as the clock at 0, which defines a universal or coordinate time T. Furthermore, the coordinate X of any point can be defined relative to 0 by the simple definition $dX/dT = \pm c$ for the motion of light.

From symmetry conditions, we must have

$$ds^2 = \exp\left[\frac{2\phi(X)}{c^2}\right](c^2 dT^2 - dX^2) - dY^2 - dZ^2. \tag{9}$$

The $g_{ij}(X)$ of Eq. (9) satisfies Einstein's field equations $R_{ij} = 0$, provided $\phi''(X) = 0$ or $\phi(X) = -gX$. The coordinate transformation

$$x = \frac{c^2}{g}\left[1 - \exp\left(-\frac{gX}{c^2}\right)\cosh\left(\frac{gT}{c}\right)\right], \qquad y = Y,$$

$$t = \frac{c}{g}\exp\left(-\frac{gX}{c^2}\right)\sinh\left(\frac{gT}{c}\right), \qquad z = Z, \tag{10}$$

reduces Eq. (9) to the simple Minkowski metric

$$ds^2 = c^2 dt^2 - (dx^2 + dy^2 + dz^2). \tag{11}$$

Thus, the coordinate system $\bar{0}$-x-y-z represents a freely falling frame of reference (inertial frame).

Returning to Eq. (9), one notes that proper time is related to coordinate time by

$$d\tau = \exp\left(-\frac{gX}{c^2}\right)dT, \qquad (12)$$

and it is precisely this fact—that the rate of an atomic clock depends essentially on the potential field ($\phi = -gX$, borrowed from Newtonian physics)—that explains the clock paradox.

It is interesting that the transformations of Eq. (10) and, hence, the metric of Eq. (9) can be obtained from the following considerations: Let $x = x_0$ be the x-coordinate of a point fixed on the freely falling frame of reference (inertial frame). Assume that the motion of this point is given by $X = X_0 + \phi(T)$, with $\phi(T)$ independent of X_0. For each X_0, there corresponds an x_0, and conversely, so that $x_0 = F(X_0)$ and $x = F[X - \phi(T)]$. Let

$$\begin{aligned} x &= F[X - \phi(T)], \\ t &= G(X, T), \end{aligned} \qquad (13)$$

with $\phi(T)$, $G(X, T)$ unknown functions. The operational definitions of T and X (namely, $dX/dT = \pm c$) for light moving in the X-direction, and the fact that $dx/dt = \pm c$ for the motion of light as observed by an inertial observer, enable one to obtain Eq. (10) and, hence, Eq. (9) (see Ref. 4).

We now consider the metric due to a perfect spherical mass. It is not difficult for observers at rest at the surface of this body to choose radial directions by means of plumb bobs and to operationally define the spherical angles θ and ϕ. Again we define coordinate time t by means of light signals sent to any observer at rest in space relative to the body (the observer can remain at rest by means of a rocket). The radial coordinate can be operationally defined by the equation $(d\rho/dt)^2 = c^2$ for the radial motion of light, but we are not certain that the coordinate speed of light normal to the radial direction is necessarily c.

Hence, the metric due to the mass must be of the form

$$ds^2 = \exp[2\lambda(\rho)](c^2dt^2 - d\rho^2) - \exp[2\mu(\rho)] \times (\rho^2 d\theta^2 + \rho^2 \sin^2\theta \, d\phi^2). \qquad (14)$$

If we compare ds^2 of Eq. (14) with the Schwarzschild metric

$$ds^2 = c^2\left(1 - \frac{2GM}{c^2r}\right)dt^2 - \left(1 - \frac{2GM}{c^2r}\right)^{-1} \times dr^2 - r^2(d\theta^2 + \sin^2\theta \, d\phi^2), \qquad (15)$$

we note that, of necessity,

$$\begin{aligned} \exp[2\lambda(\rho)] &= 1 - \frac{2GM}{c^2r}, \\ \exp[\lambda(\rho)]\,d\rho &= \left(1 - \frac{2GM}{c^2r}\right)^{-1/2} dr, \\ r &= \rho\exp[\mu(\rho)], \end{aligned} \qquad (16)$$

which yields

$$\rho = r + \frac{2GM}{c^2}\ln\frac{\left(r - \frac{2GM}{c^2}\right)}{\left(r_0 - \frac{2GM}{c^2}\right)}, \qquad r > \frac{2GM}{c^2}, \qquad (17)$$

with $\rho = r_0$ when $r = r_0$, and $\exp[\mu(\rho)] = r/\rho$, a transcendental equation in $\mu(\rho)$.

Since $\rho \approx r$ for the planets of the Sun, one would obtain the same advance in the perihelion of Mercury using the metric of Eq. (14) as well as the value for the bending of light. Moreover, the operational definition of the coordinates used in the metric of Eq. (14) precludes the idea that the advance in the perihelion of Mercury is due to an inadvertent introduction of a rotating frame of reference.

From Eq. (17), choosing the value of r_0 for the radius of the Sun, we find that ρ differs from r by approximately 15 km if we choose r as the radial "distance" of Venus from the center of the Sun. The time required for light to travel twice this distance is 10^{-4} sec. Shapiro (Ref. 5) and Muhleman and Reichley (*SPS 37-29*, Vol. IV, pp. 229–241) have suggested that the theory of relativity predicts that radar pulses sent from the Earth to Venus and back will require a slightly longer time for the journey than would be the case if space were Euclidian and light traveled with a constant speed c in straight-line motion. Their calculated difference in time is of the order of 10^{-5} to 10^{-4} sec, provided we assume that the radial coordinate of the Schwarzchild metric is identified with the astronomer's radial distance in constructing an ephemeris of the planets. However, if we use the metric of Eq. (14) with ρ defined by Eq. (17), and if we choose ρ as the astronomer's radial distance, it is easily shown that the time required for light to travel from Earth to Venus and back will differ from Muhleman's calculated time by at least 10^{-4} sec. The flaw in these previous works is that we have no way of relating a non-Euclidean radial coordinate of Einstein's theory to a Euclidean radial coordinate used in the construction of an ephemeris.

It should be noted here that the two principal checks of Einstein's theory involve angle measurements. The values for the perihelion of the planet Mercury and for the bending of light which grazes the Sun do not depend seriously on the radial coordinate involved in determining the angular momentum of the planet Mercury and on the radius of the Sun, respectively. There is no way, however, of comparing the proper time for light to travel from Earth to Venus and back with the so-called Euclidean–Newtonian time. If such experiments were to lead Shapiro, Muhleman, and Reichley to the conclusion that Einstein's theory is invalid, it is highly doubtful that relativists would take these results seriously.

References

1. Clarke, V. C., Jr., *Constants and Related Data for Use in Trajectory Calculations*, Technical Report No. 32-604, Jet Propulsion Laboratory, Pasadena, California, March 6, 1964.

2. Fock, V., *The Theory of Space, Time, and Gravitation*, Pergamon Press, New York, 1959.

3. Schilpp, P. A., *Albert Einstein: Philosopher-Scientist*, Tudor Publishing Company, New York, 1951.

4. Lass, H., *American Journal of Physics*, Vol. 31, No. 4, pp. 274–276, April 1963.

5. Shapiro, I. I., *Physical Review Letters—Fourth Test of General Relativity*, Vol. 13, No. 26, December 1964.

II. Computer Applications

A. Accuracy Studies of a Second-Sum Adams-Type Predictor–Corrector Numerical Integrator

C. J. Devine

This article is an interim report on the accuracy of a second-sum double-precision Adams-type predictor–corrector numerical integrator (subroutine FORTSZ) as a function of step size and the number of differences carried. In deciding what step size to use and how many differences to carry, one is primarily concerned with a choice that allows a certain prescribed accuracy with minimum cost (running time). The purpose of this study is to present a guide to the user of an Adams-type integrator for trajectory computation to enable him to choose optimum step size with a minimum number of differences carried. A study for circular orbits is presented here; the approach is based on that in Ref. 1, in which single-precision solutions of first-order differential equations were studied.

1. Method

In studying the errors made in a numerical solution of ordinary differential equations, the immediate problem is to decide what to use as an indication of the error made in the solution; i.e., is there a solution in closed form with which the numerical solution can be compared? In most work done in celestial mechanics, the solution is not available and the errors made in any numerical integrator are estimated.

Therefore, it was decided to start the study of a numerical integrator on a simple case, such as that for a circular orbit, and work up to the more difficult cases, such as a solution of the equations of motion describing the acceleration of a body perturbed by several other bodies or forces. In this discussion, we consider the integration of a circular orbit only.

The differential equations to be solved were

$$\ddot{x} = f(x, y) = -k^2x/r^3,$$

$$\ddot{y} = f(x, y) = -k^2y/r^3,$$

where $r^2 = x^2 + y^2$. The initial conditions were $x(0) = 1$, $y(0) = 0$, $\dot{x}(0) = 0$, $\dot{y}(0) = 1$, and $k = 1$. A solution is given by

$$x(t) = \cos kt,$$
$$y(t) = \sin kt.$$

The parameters which were varied for the different cases were the step size h and the number of backward differences of the accelerations carried.

In the second-sum Adams-type predictor–corrector numerical integrator used, the predictor and corrector formulas for the solution x are, respectively,

$$\text{P} \quad x^*_{n+1} = h^2\left(\sum_{i=-2}^{m} a_i \nabla^i\right)\ddot{x}_n = h^2\left(1\cdot\nabla^{-2} + 0\cdot\nabla^{-1} + \frac{1}{12}\nabla^0 + \frac{1}{12}\nabla^1 + \frac{19}{240}\nabla^2 + \cdots\right)\ddot{x}_n, \quad (1)$$

$$\text{C} \quad x_{n+1} = h^2\left(\sum_{i=-2}^{m} b_i \nabla^i\right)\ddot{x}^*_{n+1} = h^2\left(1\cdot\nabla^{-2} - 1\cdot\nabla^{-1} + \frac{1}{12}\nabla^0 + 0\cdot\nabla^1 - \frac{1}{240}\nabla^2 - \cdots\right)\ddot{x}^*_{n+1}, \quad (2)$$

with equivalent formulas for y_{n+1}. These formulas will be said to be of index m if the highest-order difference used is $\nabla^m\ddot{x}$.

A solution was also obtained for the first derivative of x and y. The predictor and corrector formulas for $\dot{x}$ are, respectively,

$$\text{P} \quad \dot{x}^*_{n+1} = h\left(1\cdot\nabla^{-1} + \frac{1}{2}\nabla^0 + \frac{5}{12}\nabla^1 + \frac{9}{24}\nabla^2 + \cdots\right)\ddot{x}_n, \quad (3)$$

$$\text{C} \quad \dot{x}_{n+1} = h\left(1\cdot\nabla^{-1} - \frac{1}{2}\nabla^0 - \frac{1}{12}\nabla^1 - \frac{1}{24}\nabla^2 - \cdots\right)\ddot{x}^*_{n+1}, \quad (4)$$

with similar formulas for $\dot{y}_{n+1}$. There was no occasion to use Eq. (3) in this study. The values of $\dot{x}$ and $\dot{y}$ obtained using Eq. (4) were of the same accuracy as the values of the solutions x and y and will therefore not be discussed further here.

The solution errors are computed at each integration point as the magnitudes of the differences between the computed solutions x and y and the true values of the solutions computed in extended precision (70-bit evaluation of $\sin kt$ and $\cos kt$). At $t/2\pi = 0.1, 0.2, 0.3, \cdots, p$ (every 0.1 cycle up to p cycles), the maximum error observed in each of these quantities up to t rad was output. Figs. 1–5 indicate these cumulative maximum errors for a specific number of cycles and step size and for the number of backward differences carried.

2. Analysis of Results

Truncation error is defined as the difference between the exact solution of the difference equations which approximate the differential equations and the exact solution of the differential equations themselves. Roundoff error is defined as the difference between the computed solution and the exact solution of the difference equations. The influence of truncation error is usually prominent when a large value of h is used, while the influence of roundoff error is not prominent until a greater number

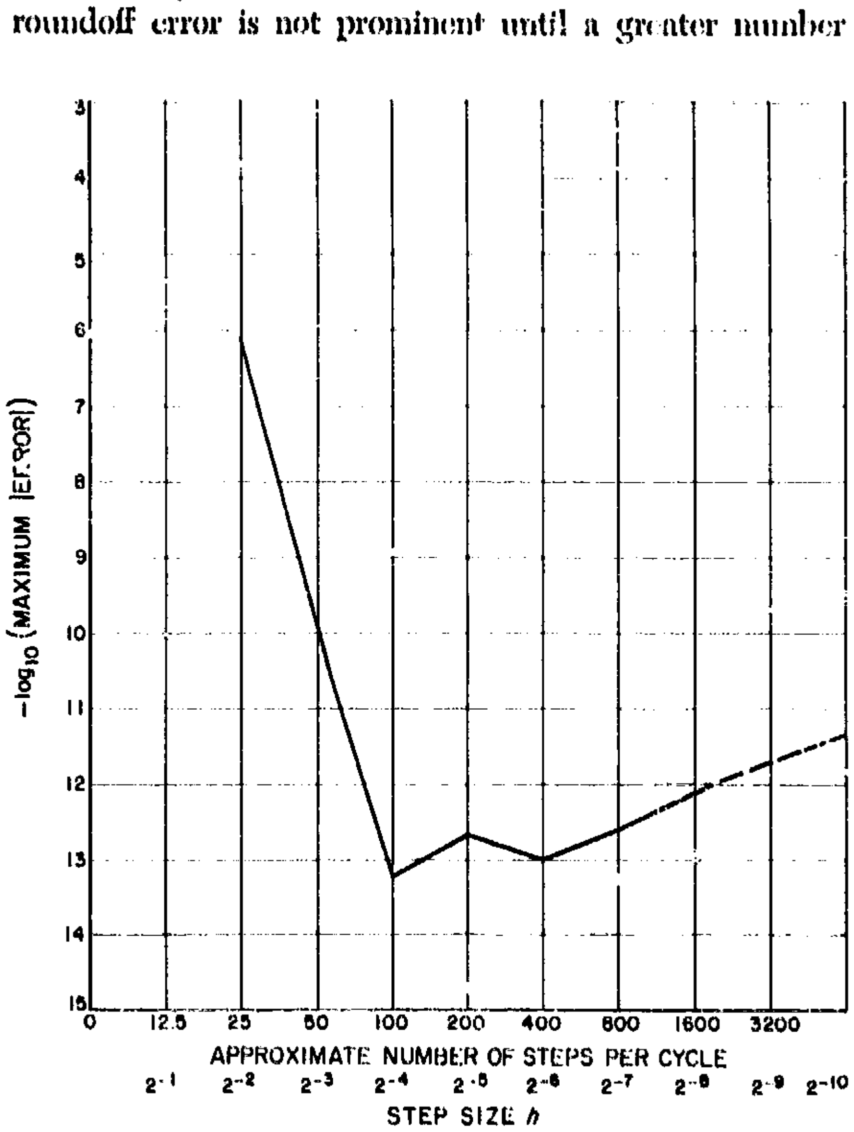


Fig. 1. Precision after 10 cycles as a function of the step size h for index $m = 9$

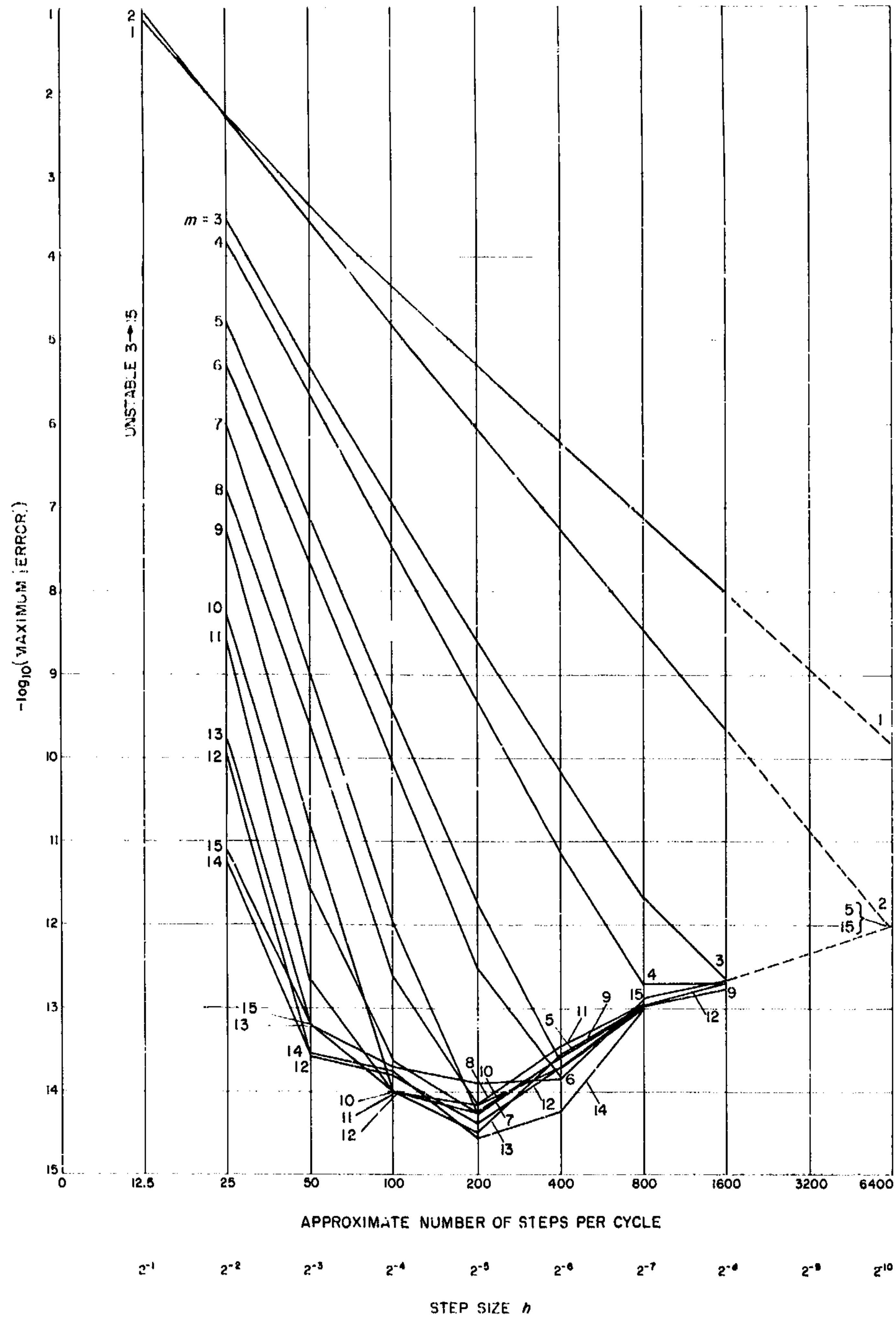


Fig. 2. Precision after 2 cycles as a function of the index m and the step size h

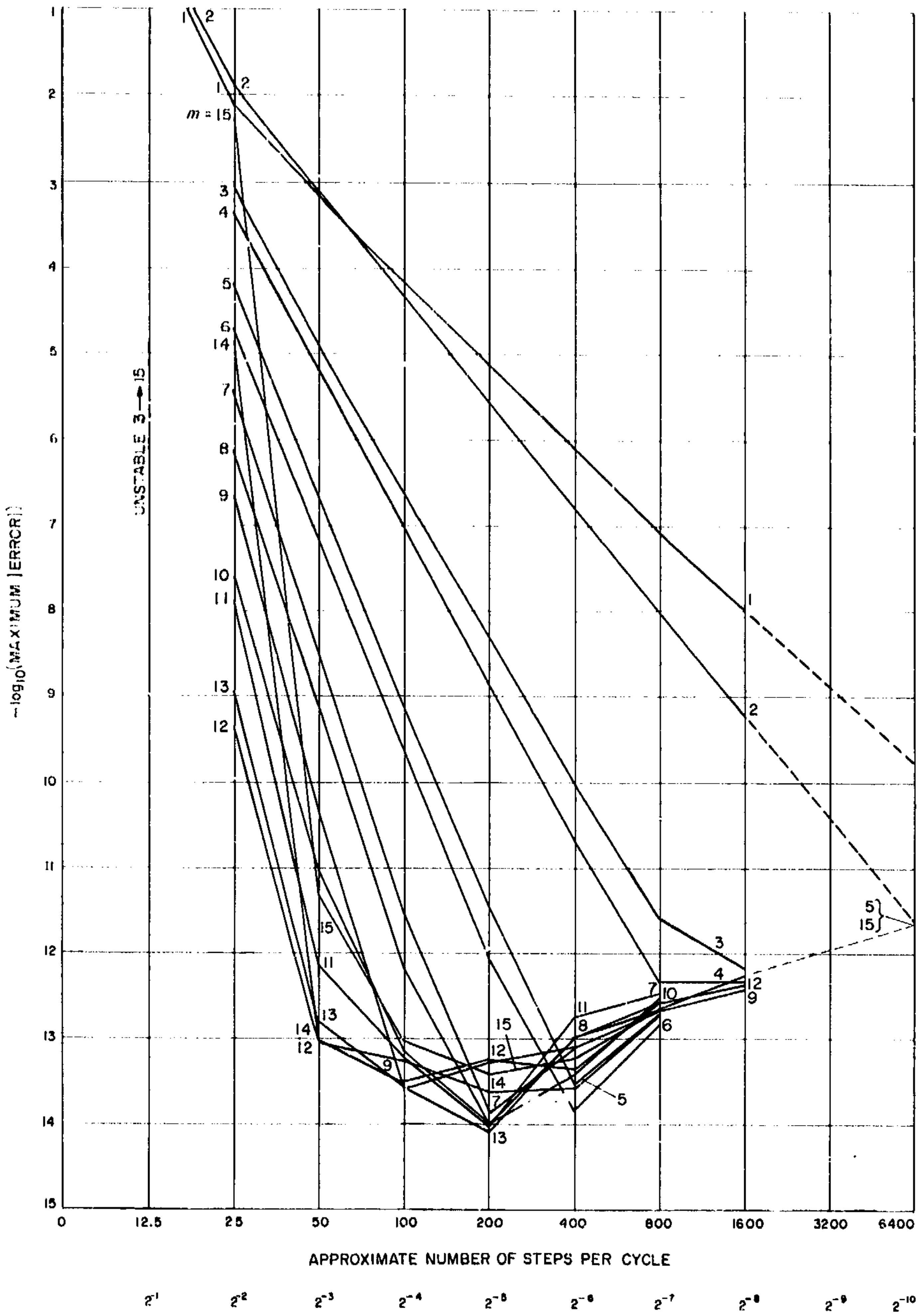


Fig. 3. Precision after 5 cycles as a function of the index m and the step size h

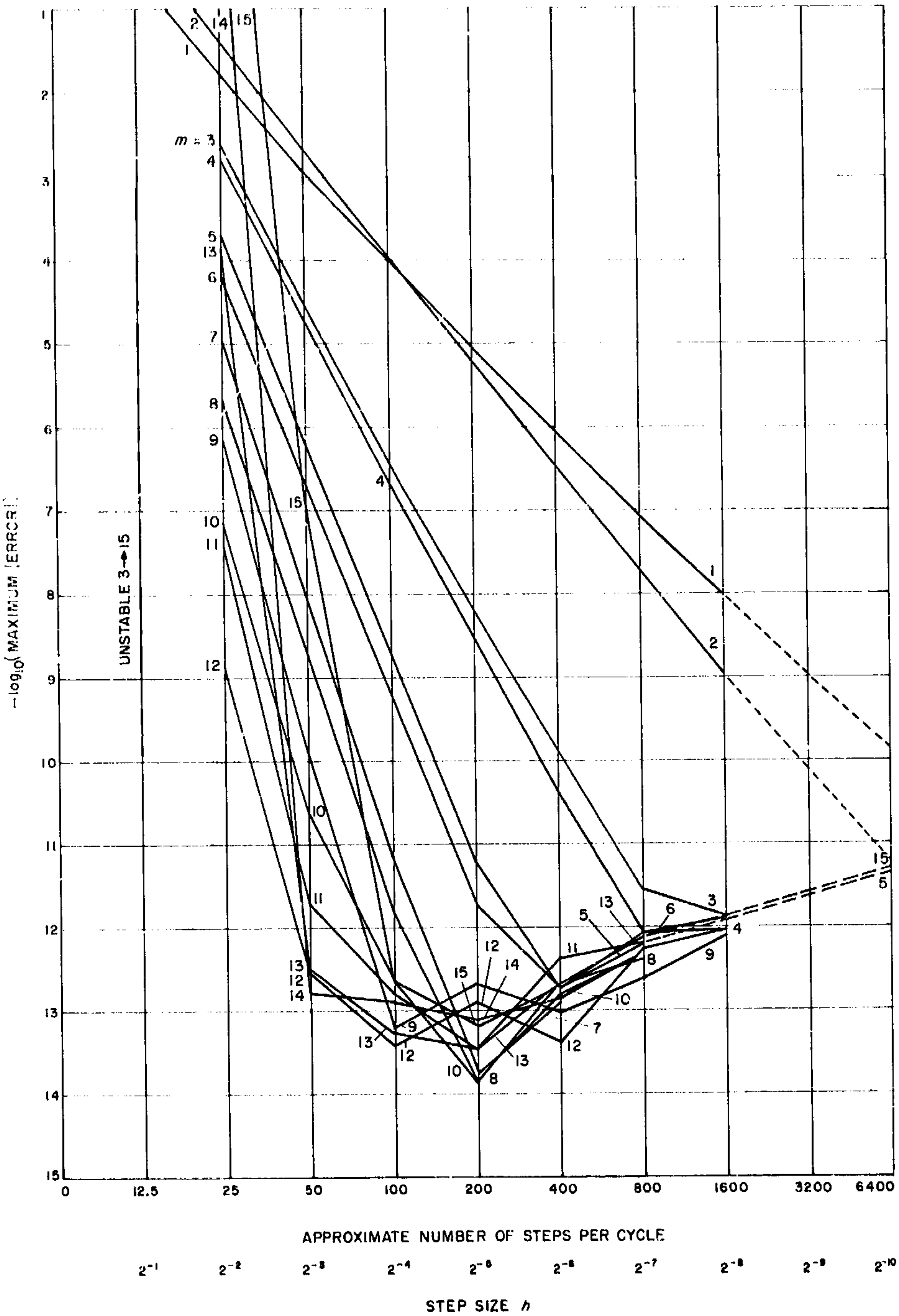


Fig. 4. Precision after 10 cycles as a function of the index m and the step size h

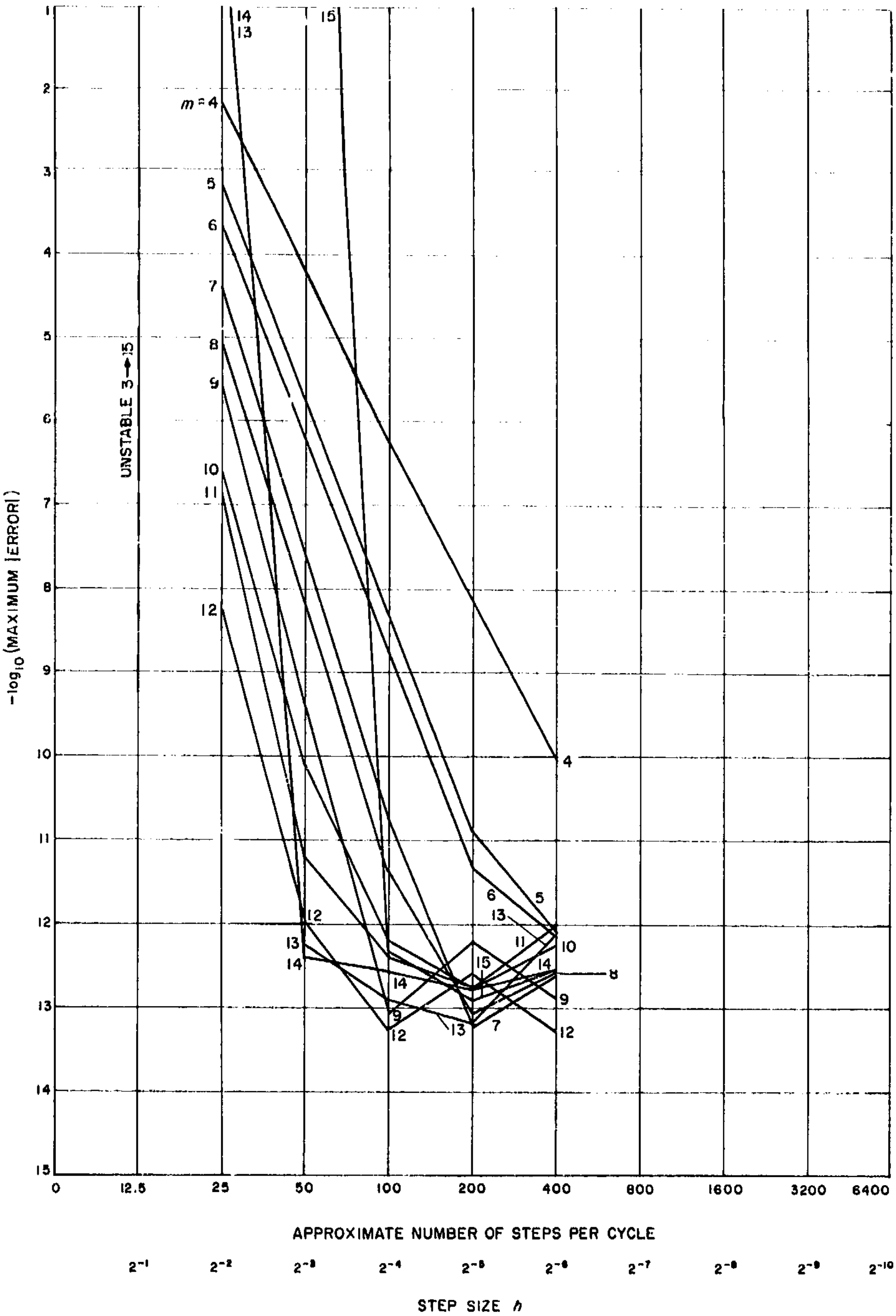


Fig. 5. Precision after 20 cycles as a function of the index m and the step size h

of steps have been taken, usually occurring when a small value of h is used. For example, a typical error curve obtained in this study is shown in Fig. 1. The predictor-corrector scheme used was index $m = 9$, and $h = 2^{-1}, 2^{-2}, \cdots, 2^{-10}$. The ordinate of Fig. 1 can be interpreted as the number of correct digits in the computed solution.

The portion of the plot where h is large represents the truncation region. Truncation error here is deterministic, and the slope of the error curve as h decreases is completely determined by the truncation error. As h continues to decrease, roundoff begins to affect the error, and a transition region is encountered. Here the error can very well attain a minimum and usually does. This may be partially explained by noting that the roundoff and truncation errors may be of equal magnitude, but of opposite signs. The curve in Fig. 1 does not represent variability in the transition region as sharply as other curves for index $m \neq 9$. As h decreases still more, roundoff error becomes dominant, and the slope is completely determined by the roundoff error; hence, this region is called the roundoff region. In all of these observations, it is assumed that we are in the stability region. For h as large as 2^{-1}, instability is immediately a problem for index $m > 2$.

Fig. 2 represents the same type of error curves for $m = 1$ to $m = 15$ for 2 cycles of integration, i.e., twice around the circle. As the number of differences retained increases (i.e., as index m increases), the number of digits of accuracy increases until a maximum is reached for $m = 14$. This is in the truncation region of Fig. 2. As h decreases, roundoff begins to influence the error, and we enter a transition region and pass on to the roundoff region where roundoff error dominates. Here, approximately the same accuracy is attained for index $m = 3$ to $m = 15$. If h decreases still further, roundoff error is completely dominant and, for all $m \geq 1$, the same accuracy is attained. It should be noted that carrying high-order differences, such as for index $m = 15$, does not in any way adversely affect the integration; however, for integrations using a value for $h \leq 2^{-8}$, it would be more economical to retain only three or four backward differences as far as storage space is concerned, with a slight decrease in running time also.

In Figs. 3–5, similar curves are represented at points further along in the integration: for 5, 10, and 20 cycles, respectively. The over-all effect is to shift all the curves upward, i.e., a gradual loss of accuracy. The instability of some of the higher-order differences is shown by the nearly vertical error slopes obtained when $h \geq 2^{-3}$. In indicating possible uses for these plots, the following remarks are pertinent: In searching for an economical way to compute a highly accurate orbit of many revolutions about a body, it was desired to ascertain whether or not a step-by-step numerical integration would be economical. In Fig. 5, for $h = 2^{-3}$ and 2^{-4} (which corresponds to approximately 50 to 100 steps/cycle), indices $m = 14$ and $m = 12$ attained the highest accuracy. Although $m = 12$ was of higher accuracy, $m = 14$ would require half the number of derivative evaluations and was therefore chosen. The integration was carried over 500 cycles and maintained an accuracy of almost 11 digits (*SPS 37-31*, Vol. IV, pp. 17, 18). As indicated by the abrupt change in the curve for $m = 14$ at $h = 2^{-3}$, the integration is as near instability as possible and yet is not in any way affected by it.

3. Conclusion and Indication of Work in Progress

The main conclusion resulting from these studies in double-precision step-by-step integration seems to be that high-order differences should be retained in the interests of economy and accuracy. They are not affected by roundoff and allow full use of double-precision arithmetic.

Similar studies over an ellipse are now in progress; all indications are that, depending on the eccentricity, the same results will follow, with a slight decrease in step size, for the same accuracy. Further study is being directed toward the economy of a predictor-only method versus that of the predictor–corrector method and the use of antecedent values of the accelerations in place of backward differences. A study is also under way to determine how satisfactory the method of integrating forward and backward is in checking the accuracy of an integration, and whether this method may be used to indicate the error when no true solution is known.

B. Computer Subroutines for Contour Plotting

C. L. Lawson, N. Block, and R. D. Garrett

In studying a function of two variables, say $z = f(x, y)$, certain properties of the function are most conveniently displayed by plotting the loci of points in the xy-plane associated with selected fixed z-values. Such loci are referred to as contours. When mechanizing contour plotting using a computer system, the problem can be separated into two independent parts: (1) the production of strings of xy-coordinate pairs representing ordered points on a contour, and (2) the translation of these strings into the code peculiar to the plotting equipment and their transmission to the plotting equipment.

Computer subroutines accomplishing (2) are in general use on the JPL IBM 7094 computers; however, a general subroutine to accomplish (1) has not previously been available. Such a subroutine, called CONTUR, has recently been completed. It was written in the FORTRAN IV programming language. The approach used in CONTUR and some of its features and limitations are described here.

1. Procedure for the Construction of Contour Strings

It is assumed that the xy region of interest is a rectangle defined by $XMIN \leq x \leq XMAX$ and $YMIN \leq y \leq YMAX$. Over this rectangle is superimposed a grid consisting of NX vertical lines at a uniform horizontal spacing of $(XMAX - XMIN)/(NX - 1)$ and NY horizontal lines at a uniform vertical spacing of $(YMAX - YMIN)/(NY - 1)$. Fig. 6 shows such a grid with $NX = 6$ and $NY = 7$. The function $f(x, y)$ is evaluated at each of the grid intersection points. The number of such points is $NX \cdot NY$. Let $Z_1, Z_2, \cdots, Z_{NZ}$ denote the Z-values for which contours are desired. For each value Z_i and each pair of horizontally adjacent or vertically adjacent grid points, a test is made to determine whether or not Z_i lies between the values of f at the two grid points. In case Z_i equals a value with which it is being compared, the decision is made as though Z_i were replaced by $Z_i + \epsilon$. When Z_i does lie between the values at adjacent grid points, a coordinate pair between these two grid points is computed by linear interpolation. Such a coordinate pair represents the approximate point of intersection of a Z_i contour with the particular grid line segment and will be called a Z_i contour point.

There remains the problem of organizing all Z_i contour points into distinct ordered strings representing ordered points on distinct contour arcs. A few simple and somewhat arbitrary rules were used.

Considering a single grid cell and a single contour value Z_i, linear interpolation assures that each of the four sides of the cell contains, at most, one Z_i contour point. The convention regarding $Z_i + \epsilon$ ensures that the number of sides containing Z_i contour points is either 0, 2, or 4. If the number is 2, the two points will be connected; if the number is 4, a further test is made and the points are connected either from Side 1 to Side 2 and Side 3 to Side 4, or from Side 1 to Side 4 and Side 2 to Side 3. (The terms and symbols used in describing an individual cell are defined in Fig. 7.)

The test is based on the assumption that f can be defined by double linear interpolation throughout the cell. In this case, its graph would be a hyperbolic paraboloid

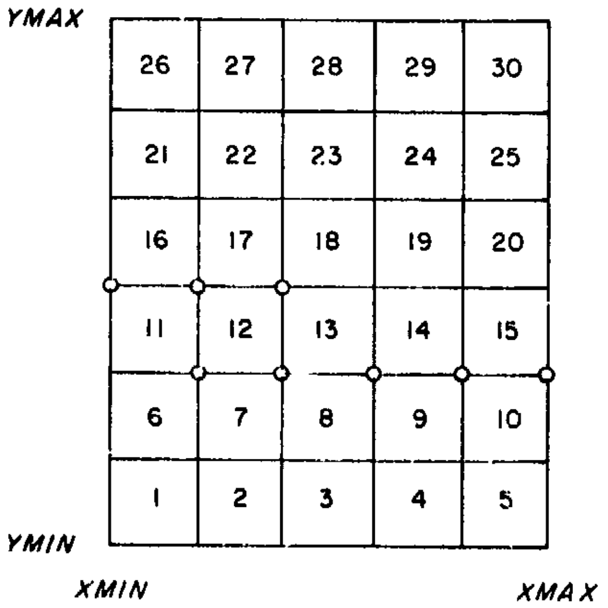


Fig. 6. Grid used by CONTUR for the case NX = 6 and NY = 7 (numbers indicate order in which grid cells are processed)

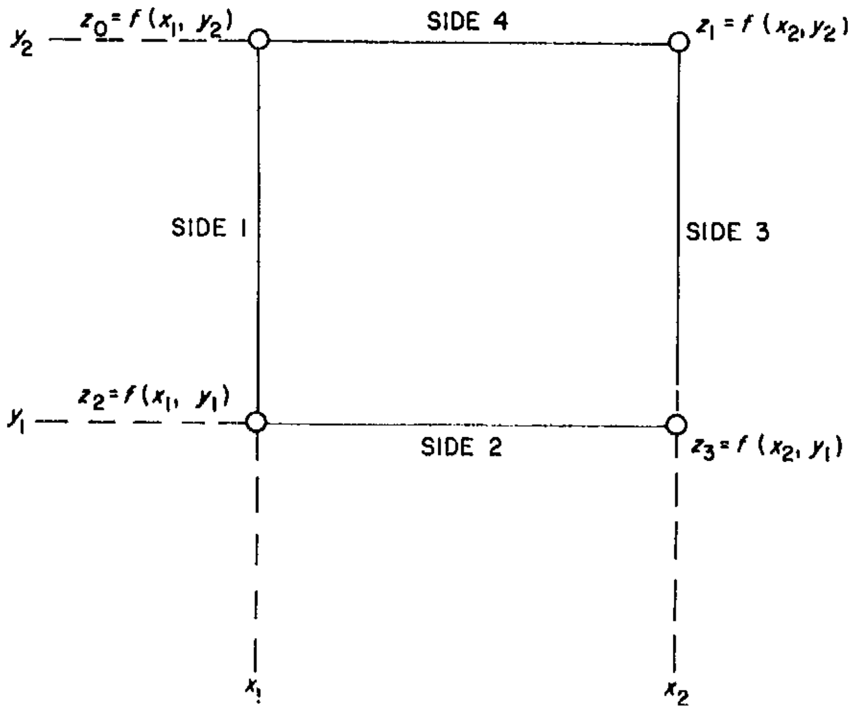


Fig. 7. Terms and symbols used in describing an individual cell

with exactly one saddle point in the open cell. The vertical and horizontal lines through this saddle point will be contour lines of the hyperbolic paraboloid. The Z_i contour points are connected so that the connecting lines do not cross these vertical and horizontal lines.

Analytically, the relative value $(0 < \xi < 1)$ of the x-coordinate of the saddle point is given by

$$\xi = \frac{(z_0 - z_2)}{(z_0 - z_2) - (z_1 - z_3)}$$

Then, if $(Z_i - z_0)/(z_1 - z_0) > \xi$, the connections are made from Side 1 to Side 2 and Side 3 to Side 4; otherwise, the connections are made in the other manner.

2. Subroutine CONTUR

In the subroutine CONTUR, the above procedure is ordered by grid cells as shown in Fig. 6. The contribution of each cell is completed for all contour values before proceeding to the next cell.

An array of $(NX + 2)$ words is used to save, temporarily, the function values computed for the upper side of each cell, so that these values will not have to be recomputed when they are later needed for the lower side of a cell in the next row. The eight grid points encircled in Fig. 6 indicate the points whose function values would be preserved while cell 12 is being processed. Note that, in progressing to cell 13, the previously saved value associated with the lower left corner of cell 12 can be discarded, and the function value for the upper right corner of cell 13 must be evaluated and saved.

The structure of the strings constructed by CONTUR will be illustrated by the following example:

Example 1: Let $f(x,y) = (x + y)^2 + 9(x - y)^2$. Let the grid be defined by $XMIN = -3$, $XMAX = 3$, $NX = 4$, $YMIN = -3$, $YMAX = 3$, and $NY = 4$. Let the desired contour values be $Z_1 = 25$ and $Z_2 = 90$. CONTUR places its results in arrays called A, B, and C. The data configuration produced by CONTUR after processing the first two cells of this example is shown in Table 1; Table 2 shows the final status of these arrays.

A triple of the form $A(J)$, $B(J)$, $C(J)$ will be called a list element with index J. The meaning of the data in Table 2 becomes clear when the list elements are arranged as in Fig. 8 and compared with the associated graphic display in Fig. 9. The minus signs on the A and B words in the arc directory lists signify that the arcs are complete, either due to having end points on the grid boundary or due to the contour in question being a closed curve.

Table 1. Contents of arrays A, B, and C after processing first two cells of Example 1

J	A (J)	B (J)	C (J)
1	25.000	3	2
2	90.000	7	0
3	6	5	0
4	−1.000	−1.875	5
5	−1.875	−1.000	0
6	0.3125	−1.000	4
7	−8	9	0
8	−0.208	−3.000	9
9	1.000	−1.964	0

Table 2. Contents of arrays A, B, and C at completion of Example 1

J	A (J)	B (J)	C (J)
1	25.000	3	2
2	90.000	12	0
3	17	−22	0
4	−1.000	−1.875	5
5	−1.875	−1.000	11
6	0.312	−1.000	4
7	−8	−18	0
8	−0.208	−3.000	9
9	1.000	−1.964	10
10	1.964	−1.000	18
11	−1.000	0.312	15
12	−13	−21	7
13	−3.000	−0.208	14
14	−1.964	1.000	19
15	−0.312	1.000	20
16	1.000	−0.312	6
17	1.875	1.000	16
18	3.000	0.208	0
19	−1.000	1.964	21
20	1.000	1.875	22
21	0.208	3.000	0
22	1.875	1.000	0

3. A Nontrivial Example

In the study of the restricted three-body problem in celestial mechanics, it is of interest to determine surfaces of zero relative velocity (Refs. 2 and 3). Applications to lunar-spacecraft-mission planning may be seen in Ref. 4. The function treated in Example 2 below is one whose

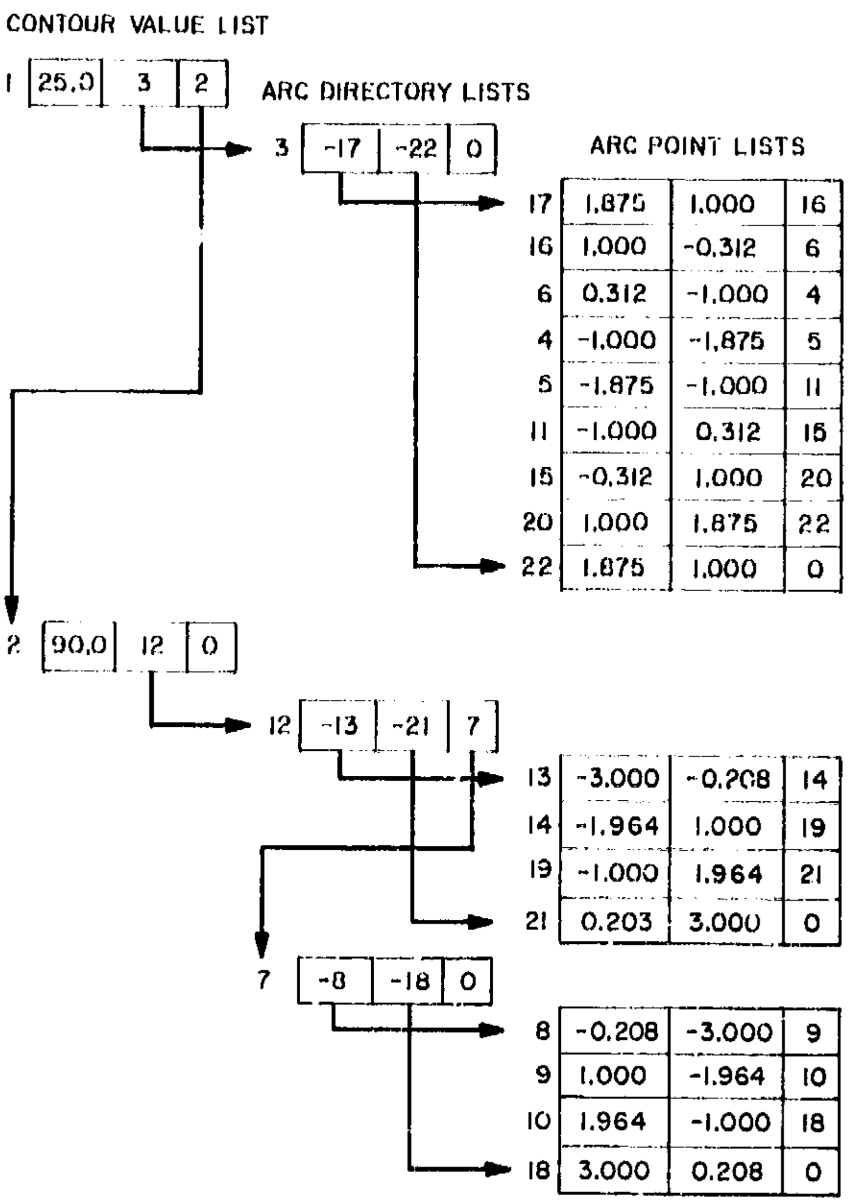


Fig. 8. List structure represented by Table 2

contours are traces of zero-relative-velocity surfaces in the xy-plane.

Example 2:

$$f(\mu; x, y) = (1-\mu)\left\{\frac{2}{[(x-\mu)^2 + y^2]^{1/2}} + (x-\mu)^2 + y^2\right\}$$
$$+ \mu\left\{\frac{2}{[(x+1-\mu)^2 + y^2]^{1/2}} + (x+1-\mu)^2 + y^2\right\}$$

where $\mu = 0.3$, $XMIN = YMIN = -2$, $XMAX = YMAX = 2$, $NX = NY = 101$, and $NF = 10$. The 10 requested contour values are given below:

A	3.050000
B	3.200000
C	3.500000
D	3.501350
E	3.600000
F	3.766413
G	4.000000
H	4.130149
I	5.000000
J	10.000000

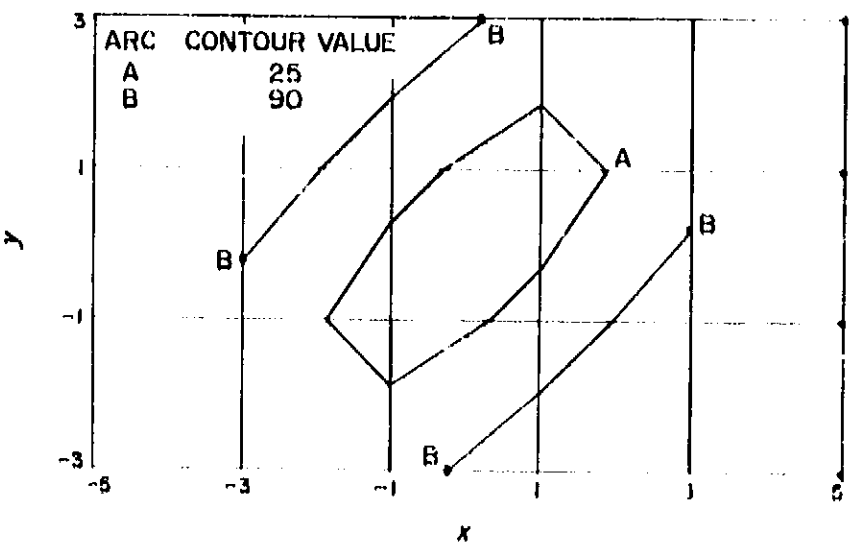


Fig. 9. Contour graph for Example 1

Fig. 10 is an unretouched reproduction of the microfilm picture produced by the Stromberg-Carlson (SC) 4020 high-speed microfilm recorder. The SC 4020 also produces the graph on an 8- × 8-in. page. For this example, the function f was evaluated 10,201 ($= 101 \times 101$) times. The total execution time for Examples 1 and 2 on the IBM 7094 computer was 1 min, 48 sec.

For Example 2, the lists constructed in the A, B, and C arrays consisted of 3460 list elements and thus occupied 10,380 words of computer storage. Space for 4000 list elements was made available to the program. If less space

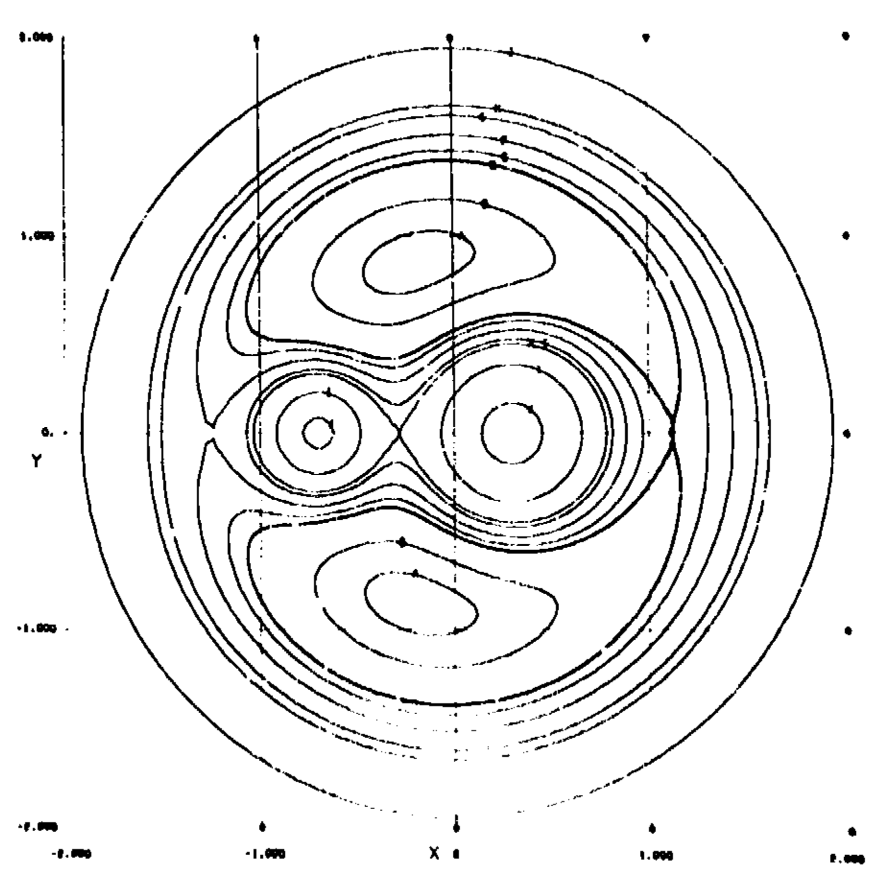


Fig. 10. Contour graph for Example 2

had been available, the program would still have completed the job, because provision has been made for the retrieval of space in the A, B, and C arrays by transmitting partially completed contour lists to the plotting subroutines.

4. Conclusions

A general-purpose FORTRAN IV subroutine for computing contour strings for arbitrary functions of two variables has been described. The subroutine has the following characteristics:

(1) It is independent of the specific plotting equipment and, therefore, can be used with basic plotting subroutines for any equipment. It could also be used as part of a computer program for purposes other than plotting, e.g., maximum seeking.

(2) Usage is simple. The new programming needed to obtain a contour graph of a new function consists of little more than the code to evaluate the function.

(3) A grid-scanning approach is used rather than a curve-following method. Resolution depends entirely on the fineness of the grid specified by the user. Since a curve-following method must include some form of scanning to avoid missing isolated curves, it is believed that the approach used here is more efficient than curve-following when graphic display is the desired end-product. For applications requiring extreme precision, the contours produced as above could be refined by gradient methods or by reapplication of CONTUR, using a finer grid on smaller regions containing the curves of interest.

(4) The number of function evaluations, $NX \times NY$, is fixed by the user's specification of NX and NY and, in particular, is independent of the number of different contour values requested.

(5) It is not necessary to be able to store the entire matrix of $NX \times NY$ grid values simultaneously. An array of $(NX + 2)$ words provides all the space needed to save values at grid points.

(6) The construction of contour strings uses list-processing techniques, so it is not necessary to anticipate the number of distinct contour curves or the number of points per curve. If the storage available for the contour strings becomes exhausted, the subroutine interrupts processing so that the strings can be sent to the basic plotting subroutines before processing is resumed.

C. Paraboloid Fitting to Minimize Path-Length Error

C. L. Lawson

In *SPS 37-31*, Vol. IV, pp. 18–21, C denotes a fixed x-y-z coordinate system. The 8-vector $\mathbf{b}$ specifies an elliptical paraboloid as follows: The first two components, b_1 and b_2, specify a rotation to obtain a new X-Y-Z coordinate system C', and the remaining components of $\mathbf{b}$ are coefficients of a polynomial function specifying an elliptical paraboloid whose axis is parallel to the Z-axis. A function $R(\mathbf{b}, \mathbf{p})$ is defined which expresses the distance measured parallel to the Z-axis between a data point $\mathbf{p}$ (represented in the C system) and the paraboloid associated with $\mathbf{b}$. The purpose here is to obtain an expression for the radiation path-length error and appropriate partial derivatives to permit the computation of paraboloid fits minimizing the sum of squares of path-length errors.

A FORTRAN II program for this purpose, described in Ref. 5, has been successfully used at JPL. The method in Ref. 5 makes essential use of the assumption that the data points being processed differ only very slightly from a paraboloid of known focal length and axial orientation. Thus, the treatment of changes of axial orientation is valid only for infinitesimal changes, and one solves for the change of focal length only after all other changes are fixed rather than simultaneously. This program has been modified by H. Christiansen of Philco Corporation to solve simultaneously for the focal length.

The computational procedure described in *SPS 37-31* and extended in this article has the following properties which distinguish it from the above-mentioned programs:

(1) The solution obtained is independent of the coordinate system in which the data are expressed.

(2) The computation is iterated until the nonlinear least-squares problem is solved, instead of accepting the solution of a linearized version of the problem.

(3) Because of (2), the procedure does not require a precise initial guess for the solution.

1. Derivation of Equations for Path-Length Error and Appropriate Derivatives

For the study of path length, it seems appropriate to fit only with circular, rather than elliptical, paraboloids so that a well-defined focus will be available. Thus, a single quadratic term of the form $b_3(X^2 + Y^2)$ will be used in place of the terms involving b_6, b_7, and b_8 in *SPS 37-31*. The formula defining the parabola in the C' system will therefore be

$$Z = b_3 + b_4 X + b_5 Y + b_9 (X^2 + Y^2) \qquad (1)$$

It will be convenient to have an alternate parameterization of the form

$$Z = a_3 + \frac{[(X - a_1)^2 + (Y - a_2)^2]}{4f} \qquad (2)$$

Here it is seen that f denotes the focal length, $(a_1, a_2, a_3)^*$ is the vertex of the paraboloid, and $(a_1, a_2, a_3 + f)^*$ is the focus. By matching coefficients of like terms in Eqs. (1) and (2), one obtains

$$a_1 = -\frac{b_4}{(2b_9)} = -2b_4 f$$

$$a_2 = -\frac{b_5}{(2b_9)} = -2b_5 f$$

$$a_3 = b_3 - \frac{b_4^2 + b_5^2}{(4b_9)} = b_3 - (b_4^2 + b_5^2) f$$

$$f = \frac{1}{(4b_9)}$$

For later use, we give the matrix of partial derivatives

$$\frac{\partial (a_1, a_2, a_3, f)}{\partial (b_3, b_4, b_5, b_9)} = \begin{bmatrix} 0 & -2f & 0 & -4a_1 f \\ 0 & 0 & -2f & -4a_2 f \\ 1 & a_1 & a_2 & a_1^2 + a_2^2 \\ 0 & 0 & 0 & -4f^2 \end{bmatrix} = E \qquad (3)$$

Let θ denote the angular separation between a data point $(X, Y, Z)^*$ and the vertex $(a_1, a_2, a_3)^*$ as subtended at the focus $(a_1, a_2, a_3 + f)^*$. The path-length error of the data point is given approximately by

$$Q = (1 + \cos \theta) R$$

where R is the Z error of the data point. [Note: The normal distance N from the paraboloid to the data point $(X, Y, Z)^*$ is related to R and Q by the equations $N = R \cos(\theta/2)$ and $Q = 2N \cos(\theta/2)$.]

Let c denote $\cos \theta$. Then c is the third component of a unit vector directed from the data point $(X, Y, Z)^*$ to the focus $(a_1, a_2, a_3 + f)^*$; i.e.,

$$c = \frac{(a_3 + f - Z)}{\rho}$$

where

$$\rho = [(a_1 - X)^2 + (a_2 - Y)^2 + (a_3 + f - Z)^2]^{1/2}$$

From $c^2\rho^2 - (a_3 + f - Z)^2 = 0$, there follows

$$c\rho^2 dc + c^2 (a_1 - X)(da_1 - dX) + c^2 (a_2 - Y)(da_2 - dY) + (c^2 - 1)(a_3 + f - Z)(da_3 + df - dZ) = 0$$

and, therefore,

$$\frac{\partial c}{\partial a_1} = -\frac{\partial c}{\partial X} = \frac{c(X - a_1)}{\rho^2} = c_1$$

$$\frac{\partial c}{\partial a_2} = -\frac{\partial c}{\partial Y} = \frac{c(Y - a_2)}{\rho^2} = c_2$$

$$\frac{\partial c}{\partial a_3} = -\frac{\partial c}{\partial Z} = \frac{\partial c}{\partial f} = \frac{(c^2 - 1)(Z - a_3 - f)}{c\rho^2} = c_3$$

2. *Organization of the Above Results for Computation*

To incorporate the above results into a computational procedure having the same structure as the procedure specified in *SPS 37-31*, it is appropriate to separate the computation into three parts as follows:

***a. Quantities depending on* b *but not on* p.** Compute the matrices **A**, $\mathbf{A}_u$, and $\mathbf{A}_p$ as in *SPS 37-31*.

$$f = \frac{1}{(4b_9)}$$

$$a_1 = -2b_4 f$$

$$a_2 = -2b_5 f$$

$$a_3 = b_3 - (b_4^2 + b_5^2) f$$

$$[e_{ij}] = \mathbf{E} = \text{matrix given in Eq. (3)}$$

b. Path-length function $Q(\mathbf{b}, \mathbf{p})$. Compute X, Y, Z, and R as in *SPS 37-31*.

$$\tilde{X} = X - a_1$$
$$\tilde{Y} = Y - a_2$$
$$\tilde{Z} = Z - a_3 - f$$
$$\rho = [\tilde{X}^2 + \tilde{Y}^2 + \tilde{Z}^2]^{1/2}$$
$$e = -\frac{\tilde{Z}}{\rho}$$
$$Q = (1 + e)R$$

c. Partial derivatives of Q. Let Q_i denote $\partial Q / \partial b_i$, $i = 1, 2, 3, 4, 5, 9$.

$$e_1 = \frac{e\tilde{X}}{\rho^2}$$
$$e_2 = \frac{e\tilde{Y}}{\rho^2}$$
$$e_3 = \frac{(e^2 - 1)\tilde{Z}}{e\rho^2}$$
$$e_0 = e + 1$$

Compute X_i, Y_i, and Z_i, $i = 1, 2$; and R_i, $i = 1, 2, 3, 4, 5, 9$, as in *SPS 37-31*.

$$Q_1 = e_0 R_1 + R(-e_1 X_1 - e_2 Y_1 - e_3 Z_1)$$
$$Q_2 = e_0 R_2 + R(-e_1 X_2 - e_2 Y_2 - e_3 Z_2)$$
$$Q_3 = e_0 R_3 + R e_3$$
$$Q_4 = e_0 R_4 + R(e_1 e_{12} + e_3 e_{32})$$
$$Q_5 = e_0 R_5 + R(e_2 e_{23} + e_3 e_{33})$$
$$Q_9 = e_0 R_9 + R[e_1 e_{11} + e_2 e_{21} + e_3(e_{31} + e_{41})]$$

3. Covariance Matrix for a_1, a_2, a_3, and f

Let $\hat{\mathbf{b}} = (b_1, b_2, b_3, b_4, b_5, b_9)$ and $\hat{\mathbf{a}} = (b_1, b_2, a_1, a_2, a_3, f)$. Let $\mathbf{T}$ denote the 6×6 matrix

$$\begin{bmatrix} \mathbf{I} & 0 \\ 0 & \mathbf{E} \end{bmatrix}$$

where

$$\mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and $\mathbf{E}$ is the 4×4 matrix of Eq. (3). Then, $d\hat{\mathbf{a}} = \mathbf{T} d\hat{\mathbf{b}}$. The covariance matrix of $d\hat{\mathbf{a}}$, and thus of $\hat{\mathbf{a}}$, is given by $\mathbf{T}[\mathrm{cov}(\hat{\mathbf{b}})]\mathbf{T}^*$. The computation of $\mathrm{cov}(\hat{\mathbf{b}})$ is specified in *SPS 37-31*.

References

1. Hull, T. E., and Creemer, A. L., "Efficiency of Predictor-Corrector Procedures," *Journal of the Association for Computing Machinery*, Vol. 10, pp. 291–301, 1963.
2. Brouwer, D., and Clemence, G. M., *Methods of Celestial Mechanics*, pp. 257–260, Academic Press, New York, 1961.
3. Moulton, F. R., *An Introduction to Celestial Mechanics*, pp. 281–294, The Macmillan Company, New York, 1914.
4. *Lunar Missions and Exploration*, pp. 7–12, Leondes, C. T., and Vance, R. W., Eds., John Wiley & Sons, Inc., New York, 1964.
5. Barondess, S. M., and Utku, S., *Computation of Weighted Root Mean Square of Path Length Changes Caused by Deformations and Imperfections of Rotational Paraboloidal Antennas*, Technical Memorandum No. 33-118, Jet Propulsion Laboratory, Pasadena, California, March 1963.

III. Environmental Requirements

A. A Microbiological Survey of Hughes Aircraft Company Facilities Involved in the Assembly and/or Testing of *Surveyor* Spacecraft

J. J. McDade, A. S. Irons, and V. J. Magistrale

Information on the levels of microbial contamination within spacecraft assembly and testing areas is not available. The study reported here represents an accelerated program designed to determine the levels of microbial contamination present within the Hughes Aircraft Company (HAC) facilities involved in the assembly and/or testing of *Surveyor* spacecraft. Emphasis was placed on the microbial contamination of surfaces in an attempt to determine the number of viable aerobic microorganisms that may accumulate on initially sterile surfaces over an 8-wk period. A limited number of air samples were collected to obtain an estimate of the number of viable particles that are rendered airborne during the entire course of a routine work day.

1. Experimental

a. Areas within HAC that were sampled. Fig. 1 is a schematic representation of the five HAC buildings sampled. As shown, 14 surface-sampling sites were selected to determine the levels of microbial contamination that may accumulate within these buildings over an 8-wk period. Eight air-sampling sites may also be identified in Fig. 1. The location and number of surface- and air-sampling sites were determined by a number of factors, including the size of the area to be sampled, the number of persons working in the area, and the type of work being done in the area.

b. Collecting (microbial-fallout-sampling) surfaces. Sterile stainless-steel strips (1×3 in.) were used as the microbial-fallout-sampling surfaces.

c. Preparation and sterilization of the stainless-steel strips to be exposed in the HAC facilities. The strips were cleaned by: (1) washing in hot tap water containing detergent, (2) rinsing with hot distilled water, (3) rinsing with isopropyl alcohol, (4) rinsing with ether, and (5) allowing to drain dry.

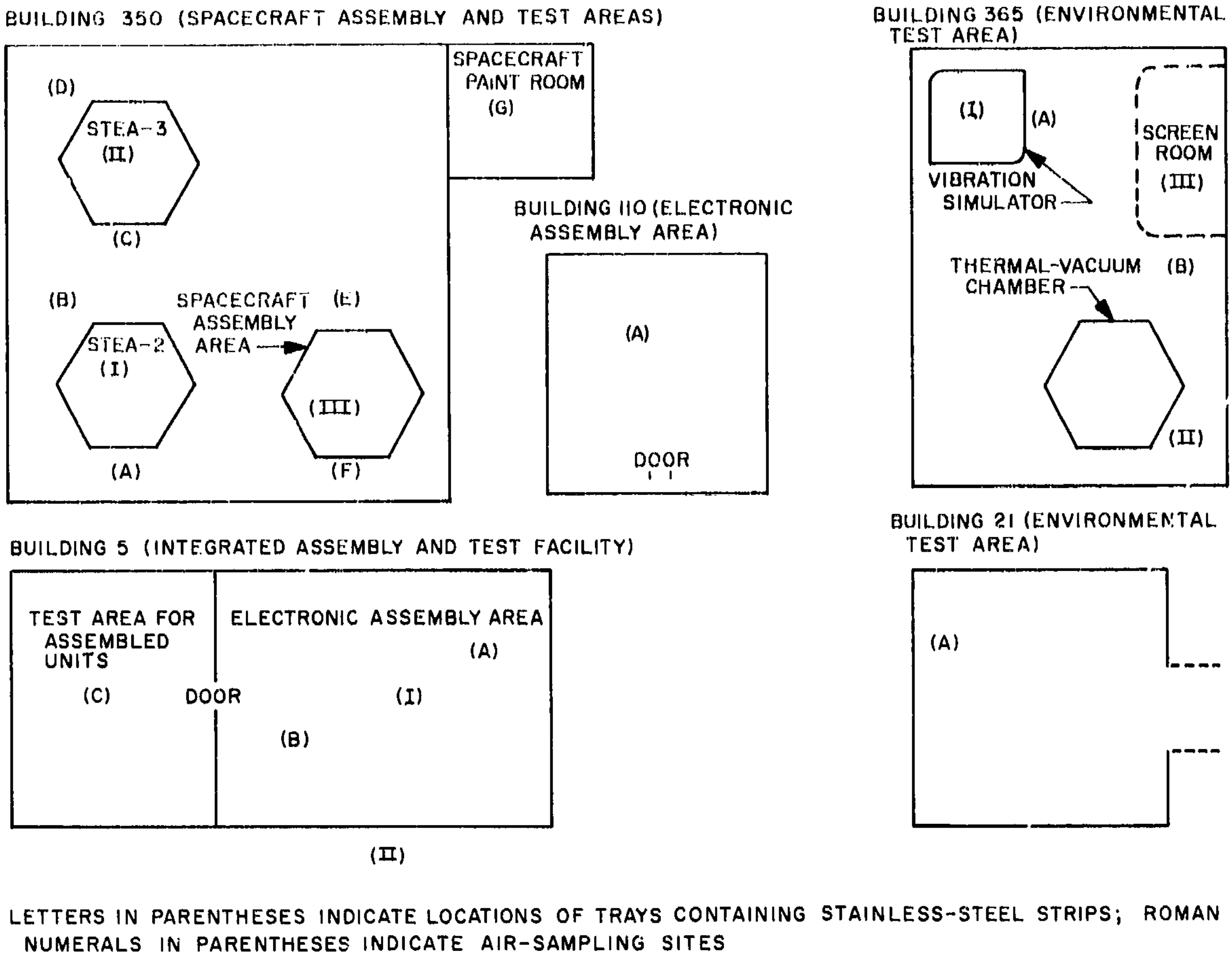


Fig. 1. Microbiological sampling sites at the Hughes Aircraft Company

Metal baking sheets (10.5 × 14.5 in.) were used as carrier trays for the strips. After cleaning, 20 strips were placed on the bottom of each carrier tray as a monolayer of strips. Fifteen trays were prepared (14 for the actual sampling and one for a sterility control check). Each was wrapped with a single layer of aluminum foil, taking care to ensure that the top and sides of the tray were covered and that the foil was not punctured in the process. The wrapped trays were sterilized by exposure to dry heat (175°C for 1.5 hr).

d. Environmental exposure of the sterile stainless-steel strips. An aluminum-foil-wrapped tray was placed in each of the 14 microbial-fallout-sampling sites shown in Fig. 1. At each site, the aluminum foil was carefully removed from the tray, and, if necessary, the sterile strips were rearranged as a monolayer with sterile forceps. Trays were placed at work height, i.e., at the same height at which *Surveyor* spacecraft hardware was being assembled and/or tested.

e. Procedure for assaying the number of viable microorganisms contained on stainless-steel strips after environmental exposure. For each microbiological assay, 5 exposed strips were randomly selected from each of the 14 microbial-fallout-sampling sites. Each strip to be assayed was placed in a separate, sterile, labeled bottle. Sterile rubber stoppers were inserted into the tops of the bottles, and the strips were hand-carried to the U. S. Public Health Service Field Station at Phoenix, Arizona. Strip pickup schedules were selected to coincide with commercial jet flights, and it was possible to begin microbiological assay of the strips 3.0 to 3.5 hr after pickup at HAC.

To assay each set of five strips from each of the 14 sampling sites, 50 ml of sterile 1.0% peptone water was added to each bottle containing an environmentally exposed strip. Next, the bottle–peptone water–strip combination was mechanically shaken for 5 min. After shaking, the five samples in each series were processed as follows:

(1) One of the five bottles in each series was put into an 80°C water bath and heat-shocked (80°C for 15 min). After heating, the liquid contents of the bottle were aseptically put into a sterile 150-mm-D Petri plate. Then, 50 ml of sterile double-strength trypticase soy agar was added to the heated liquid, and the contents of the Petri plate were carefully, but thoroughly, mixed. Following mixing, the plates were allowed to stand at room temperature until the agar had solidified (approximately 15 min).

(2) The remaining four bottles in the series were processed as in (1) above, with one exception: the heat-shocking step was eliminated; i.e., the liquid contents were plated immediately after shaking.

All plates were incubated at 32°C for 72 hr. A preliminary count was made after 40 hr of incubation, and a final count was made after 72 hr at 32°C.

f. Procedure for collecting air samples. A Reyniers slit sampler equipped with a 1-hr clock motor was used to collect volumetric air samples. All samplers were calibrated to operate at a sampling rate of 1 ft³ of air/min. The duration of each sampling period was 60 min (thereby sampling 60 ft³ of air for each sample). Trypticase soy agar was used throughout the air-sampling phase of the study. All plates were incubated at 32°C for 72 hr. After incubation, colony counts were made, and viable particle recovery per unit volume of air was calculated and recorded.

2. Results

a. Air-sampling studies. Figs. 2–5 show the results of the air-sampling studies conducted in HAC Building 350.

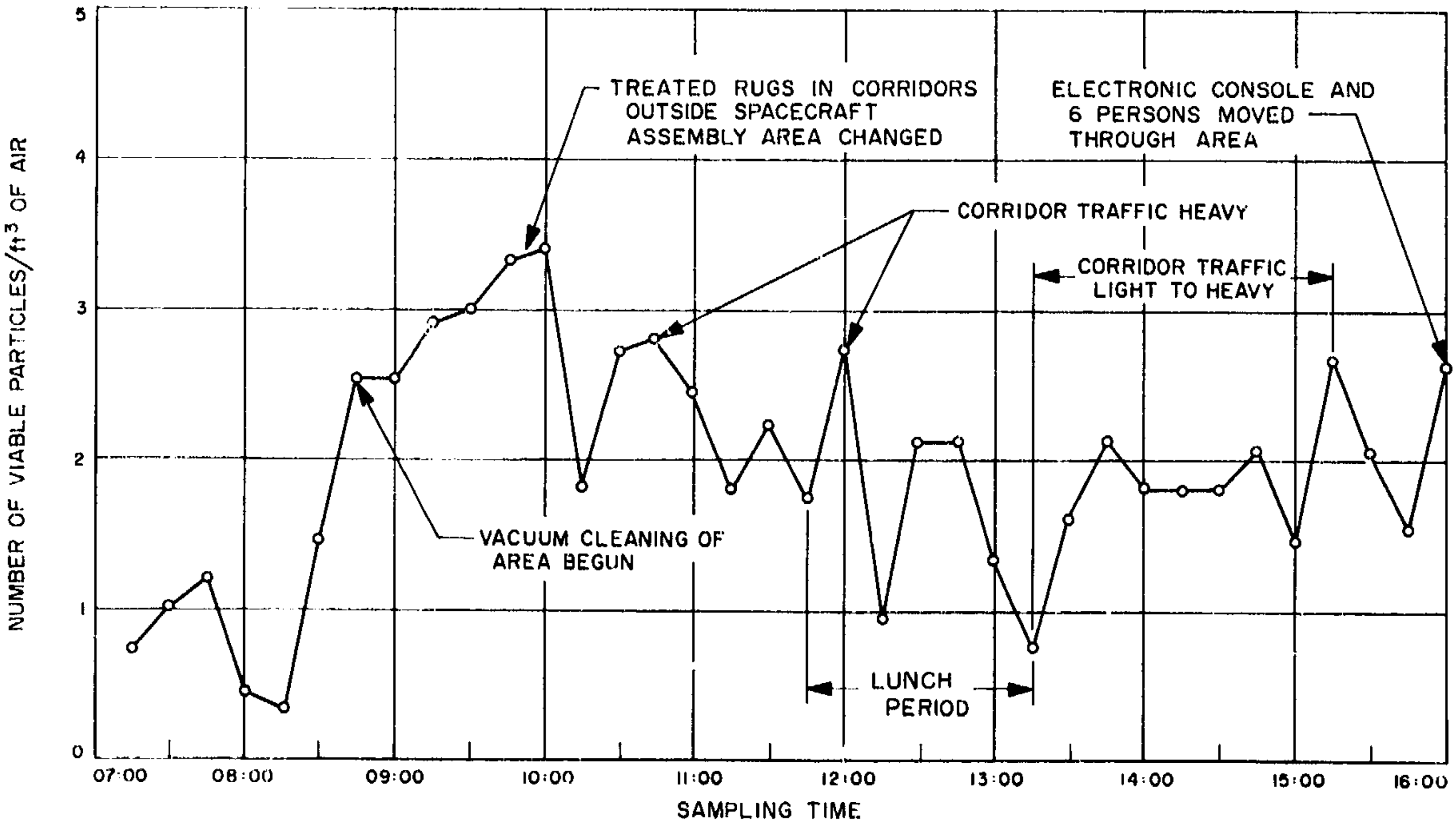


Fig. 2. Results of air sampling in spacecraft assembly area, HAC Building 350, Site III

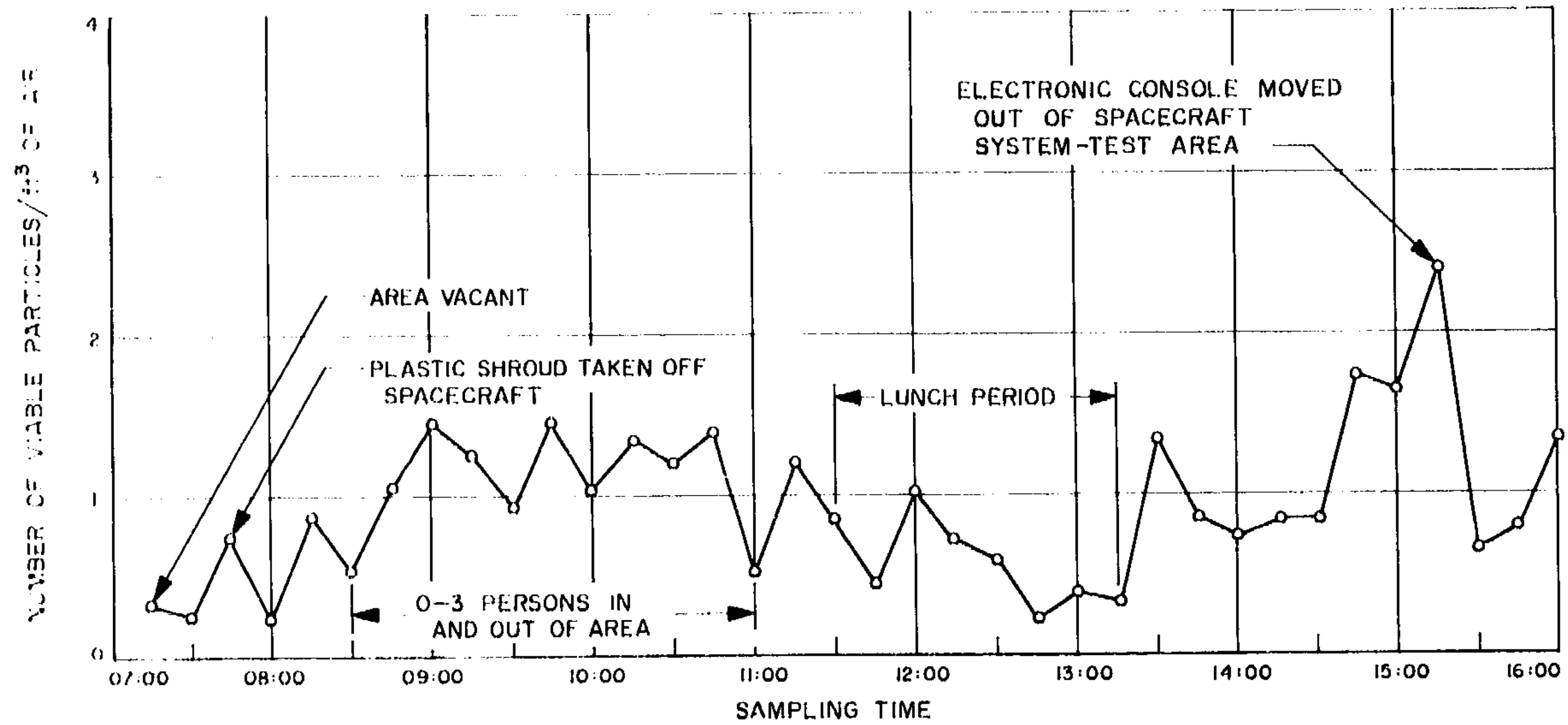


Fig. 3. Results of air sampling in STEA-2 area, HAC Building 350, Site I (little or no activity in area the entire day)

The spacecraft assembly and test area of Building 350 is a Type I controlled area.[1] As such, this area represents a floor area of 36,000 ft^2 and houses a maximum of 200 persons. Personnel in the area are required to wear caps and gowns. The area has stratified air control [(2.5 ft^3/min)/ft^2 below the 10-ft level], a temperature range of 72 $\pm$5°F, and ambient relative humidity (RH).

Fig. 2 presents the results obtained during 9 hr of continuous air sampling in the spacecraft assembly area of Building 350. *Surveyor* spacecraft hardware was not in the area. A rise in the number of airborne viable particles frequently resulted from an increase in personnel activity in the area. Vacuum cleaning, rug changing, and heavy corridor traffic all tended to increase the number of airborne viable particles. The highest number of viable particles/ft^3 of air in this area was 3.5.

Fig. 3 presents the air-sampling results obtained at the System-Test Equipment Assembly (STEA) 2 area of Building 350 during the course of a relatively inactive working day at this sampling site. Throughout the entire day, the number of airborne viable particles rarely exceeded 1.5/ft^3 of air. Near the end of the work shift, the airborne viable particle count rose to 2.5/ft^3 of air during a short period when an electronic console was moved out of the area by 5 or 6 persons. Following removal of this unit, personnel activity in the area diminished, and the number of airborne viable particles also decreased. Since there was very little activity in the STEA-2 area during this sampling day, Fig. 3 tends to serve as a representation of the "background" level of airborne microbiological contamination in this area.

The results of air sampling in the STEA-2 and STEA-3 areas during a normal working day are shown in Figs. 4 and 5, respectively. The general observation that an increase in personnel activity is followed by an increase in the number of airborne viable particles appears to apply to both of these areas. The highest number of airborne viable particles in both areas was approximately 6.5/ft^3 of air.

Fig. 6 contains the air-sampling results obtained in HAC Building 5, which is a Class I clean room.[2] The floor area is 12,056 ft^2, and 90 to 120 persons occupy the area. The following environmental conditions are maintained: temperature range of 72 to 73.5°F, 35 to 40% RH, and 10 air changes/hr. In Fig. 6, a rise in airborne viable particles may be seen as persons entered the room to start the morning work shift. Again, the number of airborne

[1]As defined in *Surveyor Project Bulletin No. 24*, Hughes Aircraft Company, November 10, 1964.

[2]According to *Standards and Guidelines for the Design and Operation of Clean Rooms and Clean Work Stations*, U.S. Air Force Technical Order 00-25-203, March 1, 1961.

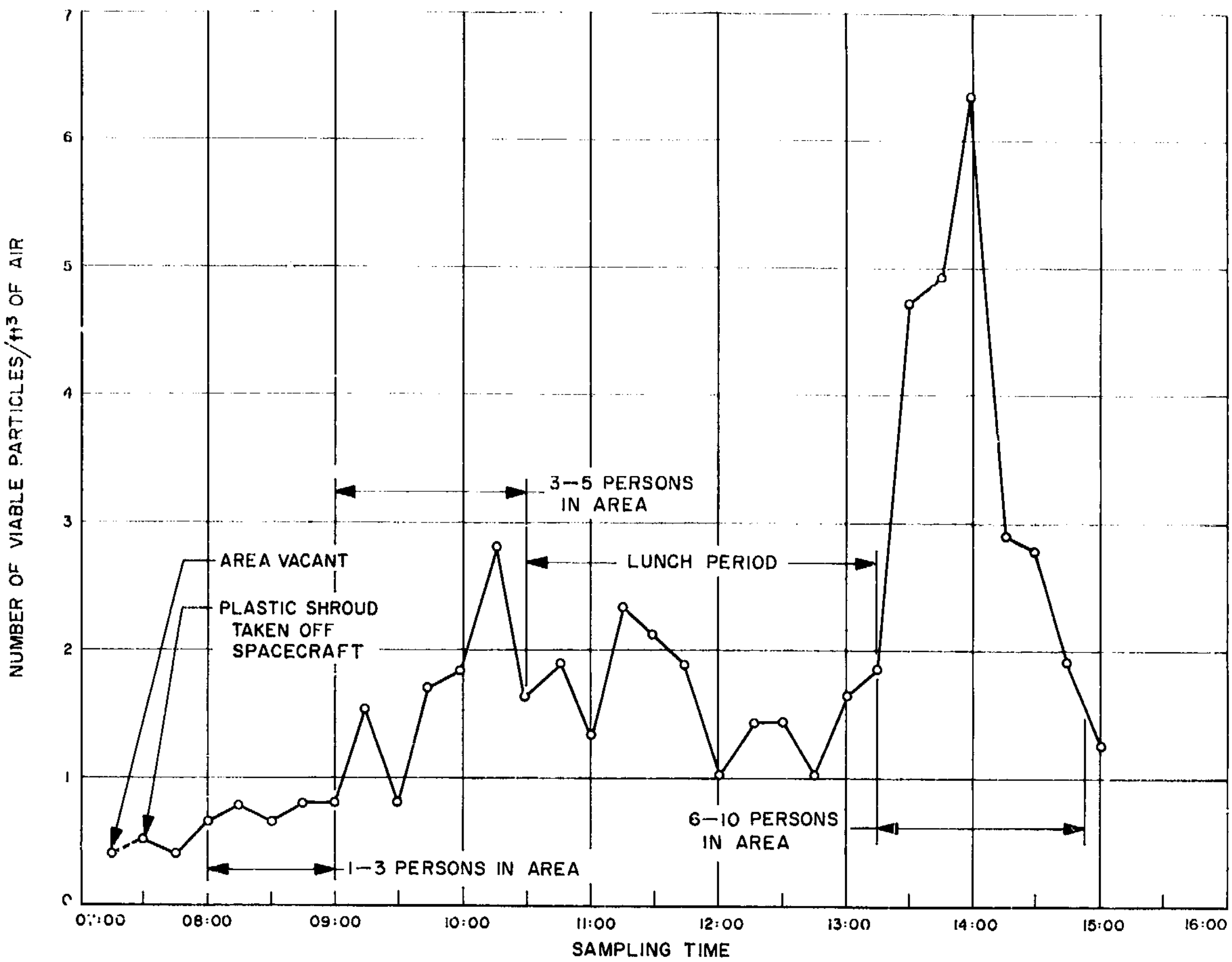


Fig. 4. Results of air sampling in STEA-2 area, HAC Building 350, Site I (area fully active; normal amount of spacecraft testing conducted)

viable particles tended to be related to the degree of personnel activity within the room.

Table 1 gives the results of air sampling in the corridor outside the integrated assembly and test facility, Building 5. The Reyniers sampler was located 5 ft from the floor and approximately 6 ft from the time clock. As may be seen in Table 1, there was a sharp increase in the number of airborne viable particles from 08:15 to 08:45. During this period, corridor activity, especially activity around the time clock, was extremely high. As corridor traffic and activity diminished, a decrease in airborne viable particles occurred.

Table 2 summarizes the results of the air-sampling studies conducted in HAC Building 365. There are two types of environmentally controlled areas in this building. The environmental test area (Sites I and II) is a Type II controlled area,[3] and the screen room (Site III) is a Type I controlled area.[3] As indicated in Table 2, the number of airborne viable particles rarely exceeded 1.0/ft^3 of air in the Type II area of this building. Air control in this area is stratified, having (4 ft^3/min)/ft^2 below the 10-ft level.

From a limited sampling of 120 ft^3 of air from Site III, it appears that airborne viable particle count never exceeded 0.3/ft^3 of air in this area. The screen room was vacant during sample collection; also, the air flow within

[3]As defined in *Surveyor Project Bulletin No. 24*, Hughes Aircraft Company, November 10, 1964.

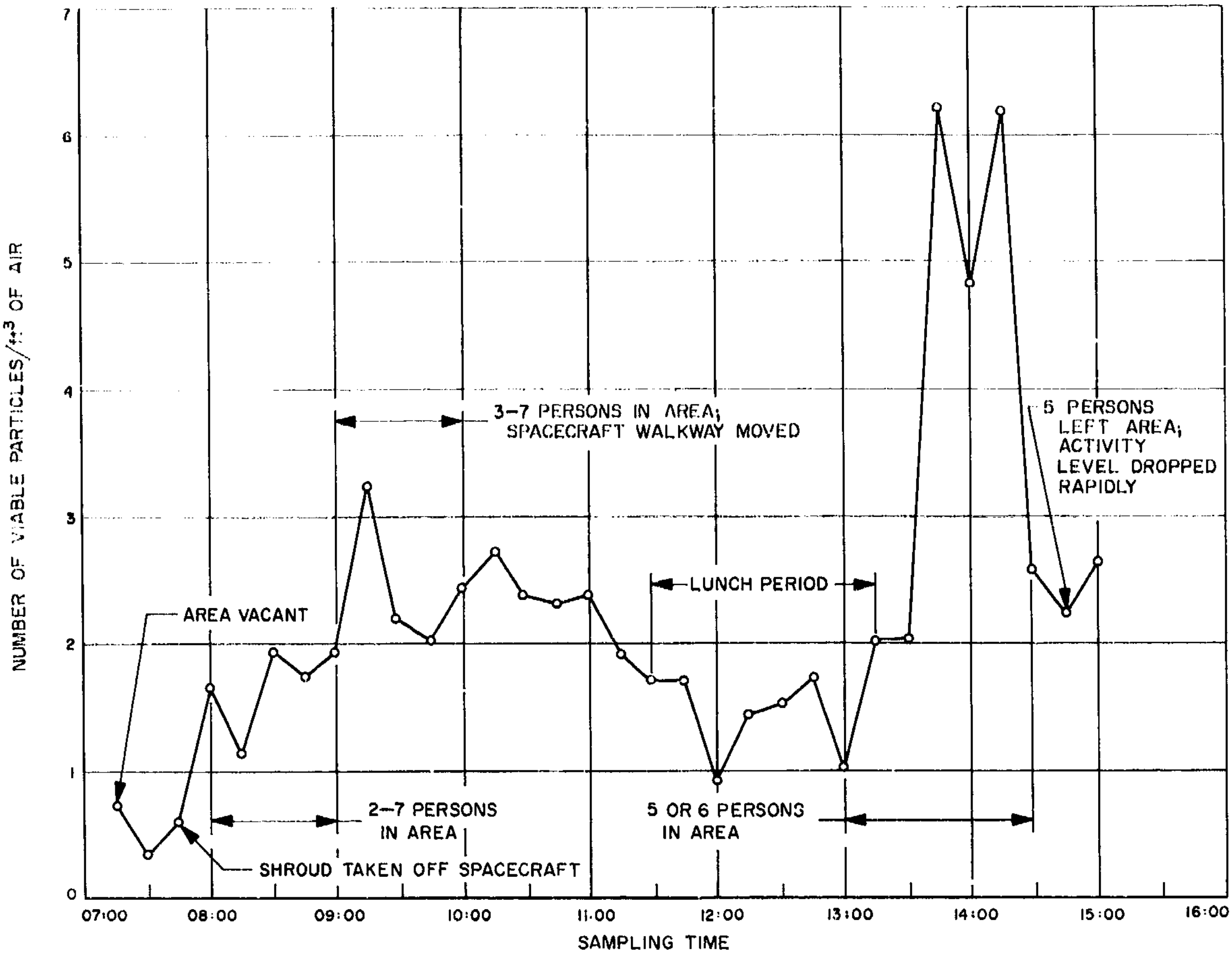


Fig. 5. Results of air sampling in STEA-3 area, HAC Building 350, Site II (area fully active; normal amount of spacecraft testing conducted)

Table 1. Results of air sampling in corridor outside integrated assembly and test facility, HAC Building 5, Site II

Sampling period	Viable particle recovery[a]		Sampling period	Viable particle recovery[a]	
	Total number of particles recovered	Average number of particles/ft³ of air		Total number of particles recovered	Average number of particles/ft³ of air
7:00–7:15	9	0.6	9:00–9:15	43	2.9
7:15–7:30	27	1.8	9:15–9:30	35	2.3
7:30–7:45	55	3.7	9:30–9:45	18	1.2
7:45–8:00	71	4.7	9:45–10:00	40	2.7
7:00–8:00 average		2.7	9:00–10:00 average		2.3
8:00–8:15	86	5.7	10:00–10:15	14	0.9
8:15–8:30	148	9.9	10:15–10:30	29	1.9
8:30–8:45	133	8.9	10:30–10:45	21	1.4
8:45–9:00	72	4.8	10:45–11:00	19	1.3
8:00–9:00 average		7.3	10:00–11:00 average		1.4

[a]Sampling rate of 1 ft³ of air/min.

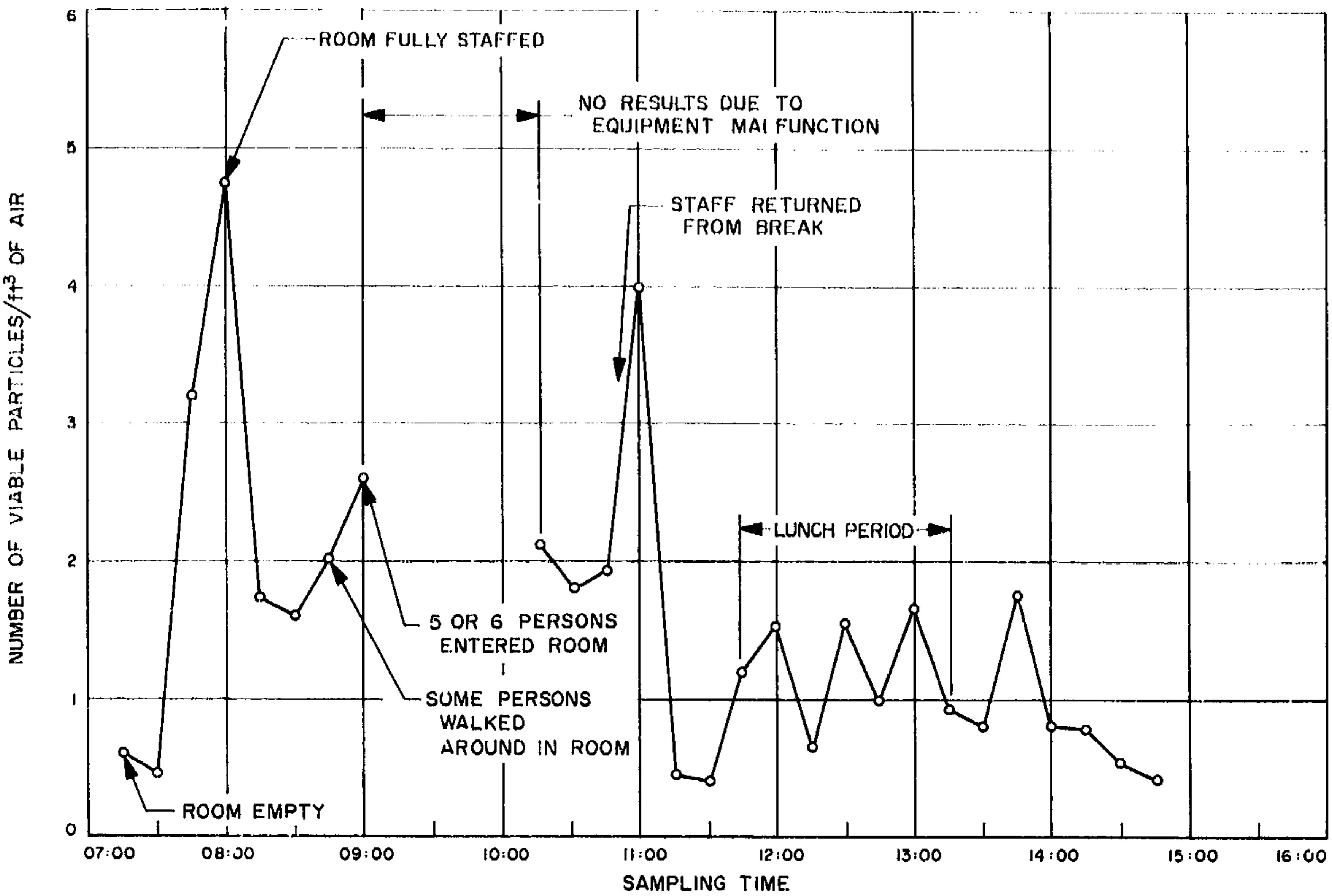


Fig. 6. Results of air sampling in integrated assembly and test facility, HAC Building 5, Site I

this area is quite turbulent. Since only a few samples were collected during a period of operational inactivity, the results described in Table 2 must be regarded as provisional. In fact, all of the air-sampling results should be regarded as preliminary results. To establish a firmer description of the level of airborne viable particles in the areas sampled, a detailed study over a period of at least 5 to 10 days of normal operational activity in all of the areas would be necessary.

b. Accumulation of microorganisms on surfaces. To conform with NASA standards,[4] the microbial recovery from the stainless-steel strips is presented in Table 3 as the number of viable aerobic mesophiles/ft² of surface. It must be realized that the actual sampling surface was a 1- × 3-in. strip, representing 3 in.² of exposed surface. A multiplication factor of 48 must be used to achieve the final expression for the number of microorganisms/ft² of surface.

Table 2. Results of air sampling in HAC Building 365

Area sampled	Duration of sampling, hr	Number of viable particles/ft³ of air	
		Range	Average
Site I	9	0.2–1.2	0.6
Site II	8	0.2–2.3	0.6
Site III	2	0.1–0.3	0.2

Frequently, considerable variation occurred in the number of viable aerobic mesophiles recovered from the stainless-steel strips collected at the same sampling site. An example of such variations is presented in Table 4.

It must be noted that this study was begun December 21, 1964. Thus, the stainless-steel strips were exposed to only 2 days of normal operational activity during the first

[4]*Interim Requirements for Bioclean Facilities*, NASA Publication (In press).

Table 3. Average number of viable aerobic mesophiles accumulating on exposed surfaces within HAC facilities[a]

Area sampled	Number of viable aerobic mesophiles/ft² of surface[b]			
	1-wk Exposure	2-wk Exposure	4-wk Exposure	8-wk Exposure
Building 5				
Site A	754	2558	1224	2160
Site B	1440	1206	2988	2400
Average	1097	1882	2106	2280
Site C	350	696	1040	3984
Building 21				
Site A	850	1570	3584	1500
Building 110				
Site A	782	998	888	—[c]
Building 350				
Site A	2352	5372	2412	3756
Site B	1152	1028	3192	30,864
Average	1752	3200	2802	17,310
Site C	802	816	1448	1956
Site D	1502	1632	1752	3408
Average	1152	1224	1620	2682
Site E	1714	1086	1920	2376
Site F	1982	974	1904	2256
Average	1848	1030	1912	2316
Average of Sites A, B, C, D, E, and F	1584	1818	2111	7436
Site G	456	678	1696	816
Building 365				
Site A	1440	1382	7760	1296
Site B	590	783	9168	2256
Average	1015	1083	8464	1776

[a]Data presented here were obtained by multiplying the average count for each 1- × 3-in. stainless-steel strip by a factor of 48.

[b]All control strips were sterile.

[c]Final results are not available for Building 110 since work in the area was stopped after the fourth week of this study. The area was vacated, and the tray of microbial-fallout-sampling strips was inadvertently removed, thereby eliminating this sampling site.

week due to the Christmas holidays; a similar restriction of normal activity was encountered during the second week of the study due to the New Year's holidays. The period from the second week to the eighth week represents a period of normal operational activity (5-day work weeks) in all of the areas in which strips were exposed.

The data in Table 3 are presented graphically in Figs. 7–10. Fig. 7 shows the number of viable aerobic mesophiles recovered after 1, 2, 4, and 8 wk of exposure in the STEA (2 and 3) areas of Building 350. As shown, the microbial contamination on strips exposed in the STEA areas seemed to stabilize within the range of 10^3 (1000 to 9999) viable aerobic mesophiles/ft² of surface. The 8-wk plot for Site B represents the recovery from a single strip; several strips exposed at this site were accidentally contaminated, thereby deleting them from the number available for the final microbiological assay. Each of the other plots represents the average value obtained from four strips. With the exception of the 8-wk plot for Site B, the number of viable aerobic mesophiles recovered on surfaces exposed within the STEA areas ranged from about 1500 to 3500/ft² of surface throughout the study. This leveling-off, or "plateau," in surface contamination has been observed by other investigators in other areas in this country (Refs. 1–3).

Table 4. Variations in the number of viable aerobic mesophiles recovered from stainless-steel strips collected at the same HAC sampling sites

Area sampled	Total number of viable aerobic mesophiles/ft² of surface			
	2-wk Exposure		4-wk Exposure	
	Range	Average[a]	Range	Average[a]
Building 5, Site A	432–5520	2558	624–1968	1224
Building 5, Site B	480–2688	1206	384–4896	2988
Building 21, Site A	960–1920	1570	2544–4608	3584
Building 365, Site A	432–2112	1382	1440–19,536	7760
Building 350, Site E	384–2016	1086	1536–2544	1920

[a]Four 1- × 3-in. stainless-steel strips were assayed in order to form the average.

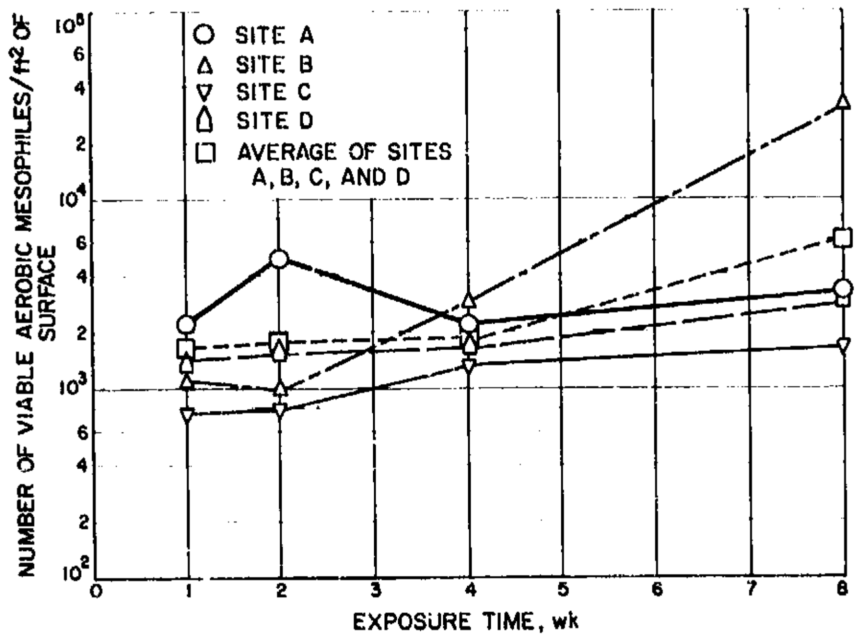


Fig. 7. Total number of viable aerobic mesophiles accumulating on surfaces exposed within the STEA-2 and STEA-3 areas, HAC Building 350, Sites A, B, C, and D

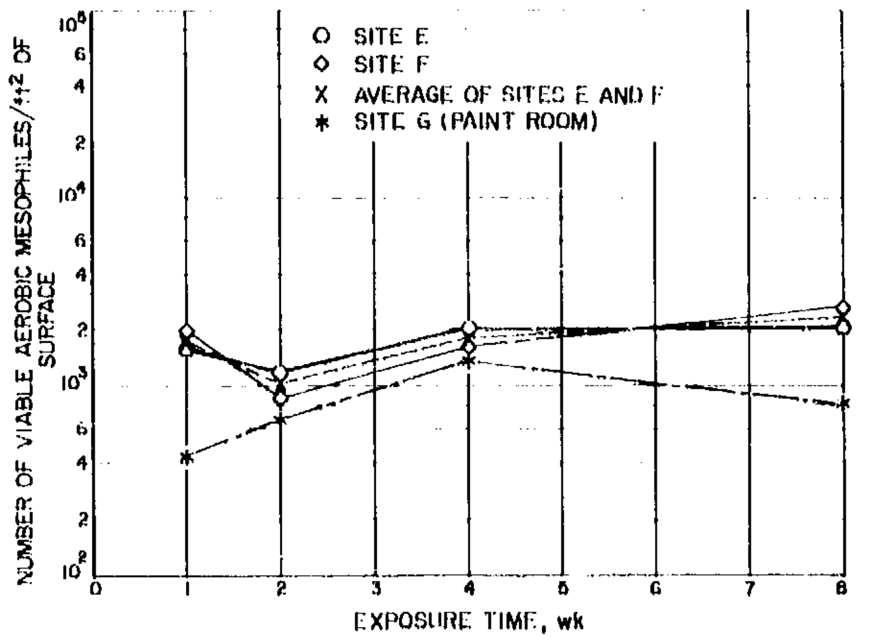


Fig. 8. Total number of viable aerobic mesophiles accumulating on surfaces exposed within the spacecraft assembly area and in the thermal paint room, HAC Building 350, Sites E, F, and G

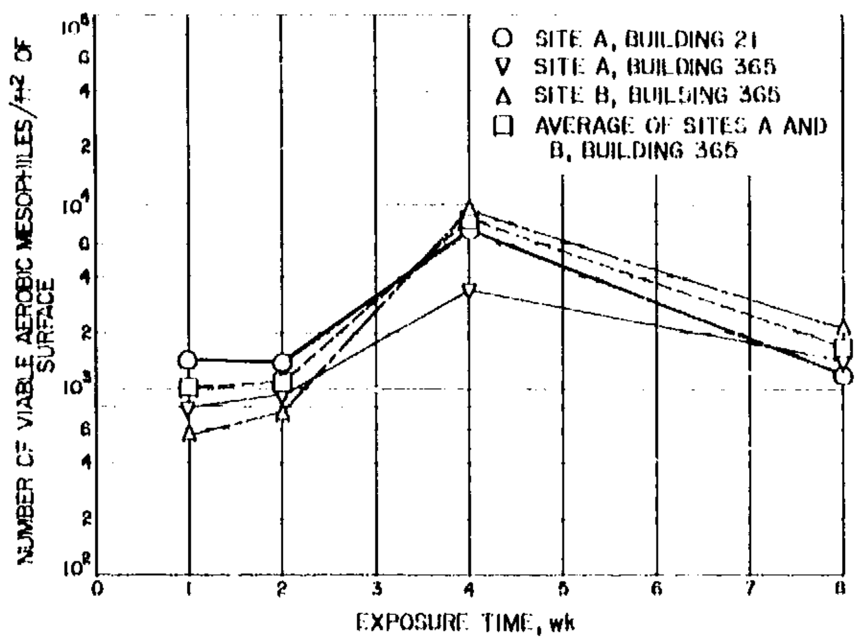


Fig. 9. Total number of viable aerobic mesophiles accumulating on surfaces exposed within the environmental test areas, HAC Buildings 21, Site A, and 365, Sites A and B

Fig. 8 shows the number of viable aerobic mesophiles recovered after 1, 2, 4, and 8 wk of exposure in the spacecraft assembly area (Sites E and F) and in the thermal paint room (Site G) of Building 350. Each plot represents the average value obtained from four strips. A *Surveyor* spacecraft was in the assembly area during the first week of the study, after which it was moved out of the area. Later, from the fourth to the eighth week of this study, another *Surveyor* spacecraft was being assembled in this area. The microbial contamination on surfaces exposed in the spacecraft assembly area again appeared to stabilize at a level of 10^3 (1000 to 9999) viable aerobic mesophiles/ft² of surface throughout the study.

Fig. 9 represents the results obtained in the environmental test areas, Buildings 21 and 365. After 4 wk of exposure, the number of viable aerobic mesophiles/ft² of surface was at a peak. From the second to the fourth week, *Surveyor* spacecraft were being environmentally tested in Building 365; however, during the fourth to eighth week, the amount of environmental testing of *Surveyor* spacecraft in this area was considerably reduced. The floor in Building 21 was being resurfaced during the 8 wk of this study; therefore, no comment shall be made on the levels of microbial contamination accumulating on surfaces exposed in this area. As seen in Fig. 9, the degree of microbial surface contamination in these areas again appeared to stabilize at a level of 10^3 viable aerobic mesophiles/ft² of surface.

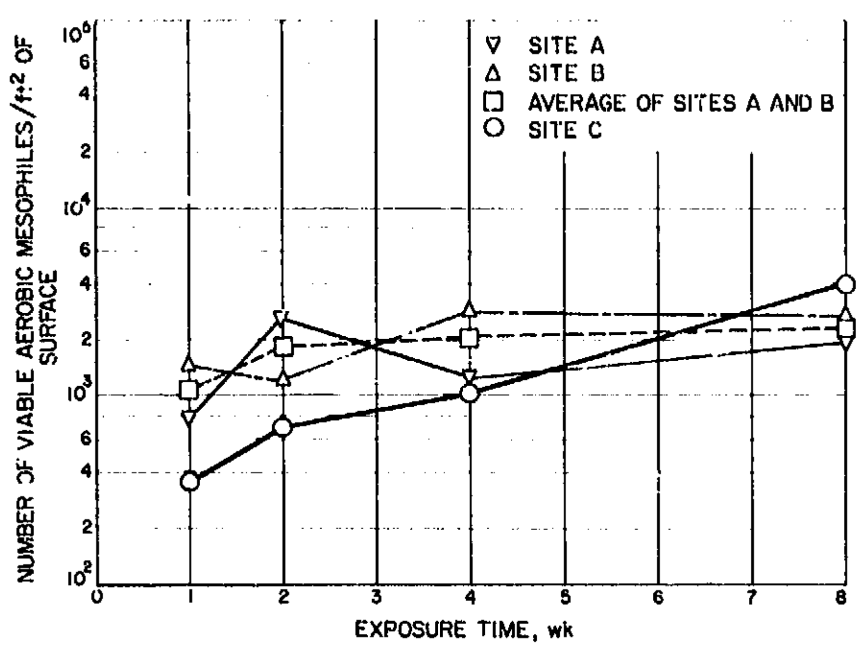


Fig. 10. Total number of viable aerobic mesophiles accumulating on surfaces exposed within the integrated assembly and test facility, HAC Building 5, Sites A, B, and C

The results obtained for surfaces exposed within Building 5 are presented in Fig. 10. Again, each plot represents the average value obtained from four stainless-steel strips. Sites A and B were located within the clean room, and Site C was located in an equipment checkout room adjacent to, but separated from, the clean room. The number of viable aerobic mesophiles/ft² of surface collected at Site C appeared to increase progressively throughout the study period. However, the number of viable aerobic mesophiles accumulating on surfaces at Sites A and B appeared to reach the range of 10^3 (1000 to 9999)/ft^2 of surface and tended to remain at that level throughout the study.

Fig. 11 presents the results obtained for the heat-shocked (80°C for 15 min) samples. Each plot represents the average results obtained from 1 (Building 21) to 6 (Buildings 5 and 365, average from 2; Building 350, average from 6) stainless-steel strips. Except for Building 21, all of the areas had a microbial spore population ranging from about 100 to 400 spores/ft^2 of surface throughout the study. However, only a limited number of samples were assayed, and, therefore, these results should be regarded as being provisional.

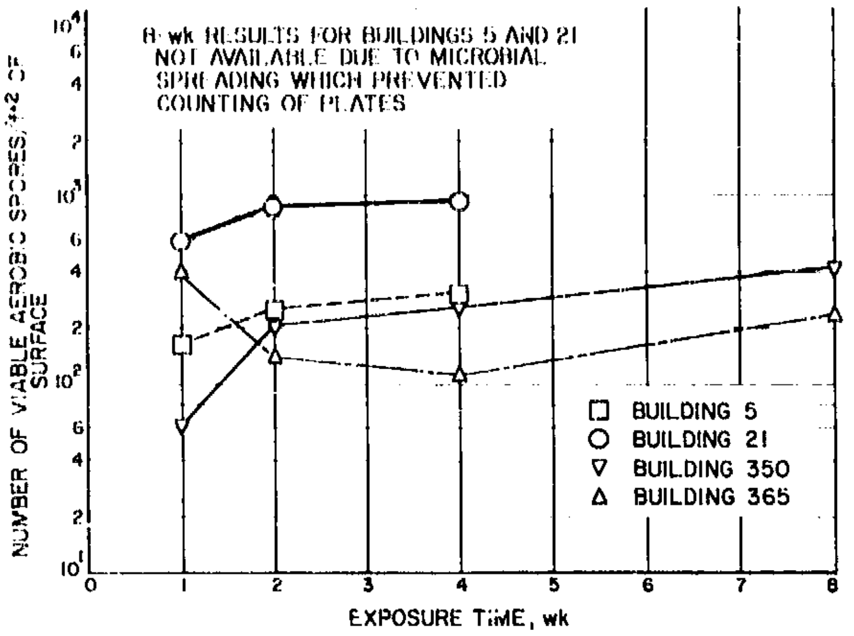


Fig. 11. Total number of viable aerobic spores accumulating on surfaces exposed within four HAC buildings involved in the assembly and/or testing of Surveyor spacecraft

3. Discussion

From the limited number of air samples collected, it would seem that there is no correlation between the air-sampling data and the surface-sampling results. The surface-sampling data (Fig. 9) indicate that Building 365, when compared with the other HAC buildings sampled, had high levels of microbial surface contamination. However, the amount of airborne viable particulate matter in Building 365 (Table 2), when compared with that in the other HAC buildings sampled, appeared to be the lowest. It must be realized that the two methods of sampling microbial contamination are different. The Reyniers slit sampler measures the number of viable particles that are suspended in air during a specific sampling interval. The stainless-steel strips measure the number of viable particles that sediment from airborne suspension, as well as the subsequent survival and accumulation of such sedimented particles. Furthermore, the methods of microbiological assay do not produce the same type of results. For example, perhaps large particles containing many viable microorganisms were shed into the environment by HAC personnel. Such particles would, if large enough, settle out of airborne suspension rapidly. If such a large particle was collected by the Reyniers sampler, only one colony would result, even though the particle contained many viable microorganisms. However, if the same particle landed on a stainless-steel strip, subsequent microbiological analysis by shaking the strip in liquid medium might fragment the large particle into clumps of smaller particles, each containing viable microorganisms. After plating and incubation, an increased number of colonies would be detected from the strip sample. Therefore, the air-sampling data are expressed as the number of viable particles/ft^3 of air, and the stainless-steel strip data are reported as the number of viable microorganisms/ft^2 of surface.

The degree of environmental contamination recovered is also determined by the microbiological sampling method used. As personnel density increases in the vicinity of a Reyniers slit sampler, "clouds" of microbiological aerosols may be generated. Collection of such aerosols would result in proportionately high levels of airborne viable particles. However, the deposition of such aerosolized particles onto stainless-steel strips might yield an entirely different set of results. Environmentally exposed microorganisms are subjected to a number of conditions that may have an effect on the survival of a given species. A number of studies (Refs. 4–13) have shown that variables such as temperature and RH, desiccation, exposure time, species, age, and number of exposed microorganisms, as well as the availability of nutrient materials, all play a role in the survival or "die-away" of microorganisms exposed to the environment. In general, the environmental conditions within most enclosed areas tend to be selective for the survival of the more resistant types of microorganisms, such as the spore-forming species. Thus, the number of viable microorganisms recoverable from the stainless-steel strips after any given exposure period is dependent upon a variety of conditions.

From the data obtained, it would appear that a plateau in the level of microbial surface contamination was reached during the course of the study. With the exception of one isolated strip, the average values for the number of microorganisms/ft^2 of surface fell within the 10^3 (1000 to 9999) range. Fluctuations in microbial surface contamination occurred within the 10^3 range, and there was suggestive evidence that, at one site (Fig. 10, Building 5, Site C), the microbial surface contamination increased progressively throughout the 8-wk study period. At present, no explanation is available for the occurrence of this plateau in microbial surface contamination. A few speculative reasons might be offered from the limited data obtained in this study; however, such speculations would require a detailed and extended study for confirmation. Perhaps the study discussed in *SPS 37-29*, Vol. IV, pp. 8–13, will provide more information in this area.

Finally, it should be mentioned that the air-sampling and spore-recovery phases of this investigation were not included in the NASA request for a microbiological survey of HAC. NASA requested JPL to obtain a "quick estimate" of the total number of aerobic microorganisms that might accumulate over a 4-wk period on initially sterile surfaces exposed within HAC facilities. However, since the first 2 wk of the study involved sample collection during atypical work weeks, it was decided by JPL to extend the study an additional 4 wk, resulting in an 8-wk study. Furthermore, the JPL sampling team expanded the microbiological assaying procedure in order to obtain some preliminary data on the microbial spore population that accumulated on the microbial-fallout-sampling surfaces. The JPL sampling team also collected a limited number of air samples in an effort to obtain an estimate of the number of airborne viable particles generated within several of the HAC facilities during the entire course of a routine work day.

Due to a multiplicity of factors, the study was conducted along the lines of a survey, and, therefore, only a limited number of microbiological samples could be collected and/or assayed. Thus, the data obtained in this study must be regarded as preliminary information. Yet, the results obtained appear to successfully terminate the investigation. Other useful information, in addition to that sought by NASA, was obtained during the study. Taken collectively, the results from this survey might also be used as guidelines to develop a future microbiological sampling program if a need for more definitive results should arise.

4. Summary

The study reported here represented an accelerated program initiated by JPL at the request of NASA to obtain an estimate of the levels of microbial contamination present within the HAC facilities involved in the assembly and/or testing of *Surveyor* spacecraft. Useful information, in addition to that requested by NASA, was obtained and should serve as guidelines for future microbiological surveys.

Due to a number of factors (e.g., urgency of need for the results; temporary lack of microbiological facilities at JPL; difficulties in transporting a large number of samples to the U. S. Public Health Service Field Station at Phoenix, Arizona; and personnel limitations), only a limited number of microbiological samples were collected. Therefore, the results obtained in this study must be regarded as preliminary information. Definite conclusions cannot be drawn from the data obtained; however, a few general statements can be made concerning the surface-sampling phase of the study:

(1) The total number of viable aerobic mesophiles recoverable from test surfaces appeared to stabilize and fall within the range of 10^3 (1000 to 9999)/ft^2 of surface over the study period.

(2) The total number of viable aerobic spore-forming mesophiles recoverable from a limited number of the test surfaces appeared to stabilize and fall within the range of 10^2 (100 to 999)/ft^2 of surface over the study period.

The number of airborne viable particles appeared to be dependent, at least in part, on the number of personnel and type of activity occurring within the area. However, no conclusions should be made from the very limited amount of air-sampling data obtained during this survey.

B. Microbiological Analysis Techniques for Spacecraft Sterilization

L. L. Reed

The ability to demonstrate the presence or absence of viable (living) microorganisms is basic to the development and specification of sterilization and decontamination procedures. Demonstrating the presence of viable microorganisms when they are present in large numbers is a relatively easy task. However, the ease of demonstration decreases as the numbers of viable microbial cells decrease; therefore, proving the presence of small numbers of living cells becomes an extremely difficult task, and demonstrating the complete absence of life is perhaps impossible, at least by using the present state-of-the-art sterility tests.

Sterility testing procedures currently used are based upon the growth response of microorganisms when placed in an environment of growth conditions assumed to be satisfactory for viable cells. There are many perplexing problems, both physical and biological, that complicate sterility test procedures. One of the greatest problems is freeing the microbial cells from interfering materials so that they are given the maximum opportunity to utilize the growth environment. This is an especially perplexing problem when attempting to analyze the microbiological content of solid spacecraft parts and materials. For nearly all solid materials, some method of diminution or pulverization is required to free entrapped microbial cells. In such methods, the size of the resulting particles becomes an important factor in determining the number of cells to be exposed to the growth environment.

Past investigations on the microbiological analysis of spacecraft parts have utilized a number of pulverizing techniques employing such implements as a hammer, pliers, drill, mortar and pestle, and blender (Ref. 14). Microbiological results were reported on the basis of growth or no growth from the cultured pulverized material. It is unfortunate that many individuals interpreted the "no growth" report as sterility. It is obvious that the methods used produced a wide variety of particle shapes and sizes, indicating the distinct possibility that the particles could still harbor many viable cells which have average diameters of the order of 0.5 to 1.0 μ.

There are no standardized procedures for the microbiological evaluation of spacecraft parts and materials, and the use of methods such as those mentioned above leaves considerable doubt as to the correlation of data derived from different studies and, in fact, within a given study. If meaningful data are to be generated for use in developing spacecraft sterilization procedures, it is necessary to employ reliable standardized microbiological evaluation procedures. A sterility test which would provide an absolute answer concerning the sterility of solid materials should: (1) provide maximum opportunity for every viable cell to demonstrate, in some manner, its viability; and (2) negate ambiguity due to various physical and biological factors. The practical sterility tests in use today do not completely satisfy the requirements of the ideal test, and it appears unlikely, at present, that a practical test will completely recover all viable microbial contamination from the interior of spacecraft solids. However, it is possible to improve the techniques by including some degree of quantitation so that numerical values can be assigned to levels of viable microbial recovery. The knowledge obtained from an investigation of feasible methods for recovering viable cells from the interior of solids should form a microbiological basis for: (1) developing adequate research techniques for the spacecraft sterilization research program; (2) developing required procedures for the clean assembly and sterilization of spacecraft; and (3) establishing reliable microbiological assay techniques for monitoring the clean assembly and sterilization of spacecraft.

The Dynamic Science Corporation of South Pasadena, California, under contract to JPL, has been investigating and evaluating various methods that might be applicable for the detection of small numbers of viable microbial cells in the interior of solids. The investigation included not only the evaluation of possible detection methods, but also the evaluation of commonly used comminution methods that would free the microbial cell from its encasement in a solid. The following discussion constitutes a terminal report of the completed contract work. The investigation consisted of the following four study phases: (1) methods for the detection of viable microorganisms in solid materials, (2) methods for freeing viable microorganisms from solid materials, (3) applicability of drilling and sectioning for freeing microorganisms from encasement, and (4) reliability of drilling and cultural methods for detecting viable microorganisms in solid materials. This article covers Phases 1 and 2.

1. Methods for the Detection of Viable Microorganisms in Solid Materials

The objective of this study phase was to examine feasible detection methods which will distinguish between living and dead cells, detect small numbers of viable microorganisms, and perhaps have an added advantage of rapid detection. The following methods were investigated:

a. Counting of stained cells. In the study of staining procedures, attention was focused on those techniques which appeared most likely to meet the requirement of being sufficiently specific to differentiate, without question, between viable and non-viable cells when only small numbers of cells are present (Ref. 15). Staining procedures may be divided into fluorescent techniques (Refs. 15, 16, and 27), which utilize ultraviolet light to observe fluorescence, and non-fluorescent techniques (Ref. 22), which use visible light to observe the color of the stained cells. The non-fluorescent methods do not appear to be sufficiently specific to always differentiate between viable and non-viable spores and vegetative or non-sporous cells. The fluorescent techniques appear to have a greater sensitivity, and, therefore, the study of staining was centered on procedures using the following: dimethylchlorotetracycline, primuline, acridine-orange, and auramine-O.

Spore suspensions of approximately 10^9 spores/ml were used in the staining techniques and were prepared from dry spore preparations of *Bacillus subtilis* var. *niger*. Both viable and non-viable spores were used. The non-viable spore suspensions were prepared by exposing viable spores to ethylene oxide, steam autoclaving, dry heat, and chlorine gas. The spores were prepared on glass slides, stained according to standardized techniques, and observed by ultraviolet fluorescent microscopy.

The results of the fluorescent staining study are summarized in Table 5. It appears that there is a slight increase in fluorescence intensity of dry-heat-killed spores when compared to viable spores or spores killed by other means. However, the differential staining reactions of the four procedures are not significant enough for practical application in sterility testing. Although these techniques did not allow adequate distinction between viable and non-viable contamination inside solids, detection of contamination inside solids is possible, and it appears feasible that staining techniques, used in combination with other sterility test techniques such as culturing, can increase the sensitivity for detecting contamination.

Table 5. Relative fluorescence of *Bacillus subtilis* var. *niger* spores treated with fluorescent dyes

Spore suspension	Relative fluorescence[a] Dimethylchlorotetracycline used	Primuline used	Acridine-orange used	Auramine-O used
Viable	+++	++	+	+++
Non-viable Steam-autoclaved	+++	—	+	+++
Exposed to dry heat	++++	+++	++	+++
Exposed to chlorine gas	+++	—	+	+++
Exposed to ethylene oxide	+++	—	+	+++

[a]Fluorescence intensity was judged semiquantitatively as + through ++++.

b. Measurement of the electrophoretic mobility of cells. The second method examined was that of electrophoresis, or the migration of fluid-suspended particles under the influence of an electric field (Refs. 17, 19, 21, and 23). When particles with an electric charge are placed in an electric field, they will move toward one or the other of the poles, depending upon the net charge of the particle. Particles with different net charges will move with different velocities; thus, separation of particles can be attained.

The electrical-charge-to-mass ratio of cells is affected by life processes, and the metabolic processes within a living cell maintain an electrical charge on the cell of a characteristic type and magnitude. Presumably, the charge on a non-viable or dead cell is different from that on a viable cell; thus, when populations of both living and dead cells are suspended in a fluid in an electrical field, the two types of cells will exhibit different mobilities. The ratio of the mobilities of these two types of cells is usually measured. The distance through which a particular particle moves under the influence of an electric current is proportional to the time and to the potential gradient at the particle's location. The potential gradient, in turn, depends upon the current, the conductivity of the solution, and the cross-sectional area of the solution at this point. The proportionality constant is known as the electrophoretic mobility.

A rectangular microelectrophoresis cell was used in this study, and the mobility of the microbial cells was observed with a microscope with a ×200 to ×1000 magnification. Using a calibrated ocular micrometer, migration velocities were determined by focusing on a given cell and measuring the time required for it to traverse a given distance. In an electrophoresis cell of the type used, the actual velocity of the microbial cell will be composed of two components: One is the result of the actual electrophoretic forces on the cell, while the other is due to the fluid motion within the electrophoresis cell. This latter motion, termed the electro-osmotic streaming effect, results from the effect of the electrical field on the suspending fluid. For a closed cell such as an electrophoresis cell, the velocity due to electro-osmotic streaming is a maximum on the centerline in one direction and is in the opposite direction along the walls.

The streaming velocity is zero at some plane between the centerline and the wall of the chamber or cell. This is the stationary plane, and the microscope must be focused in it since any motion of particles along this plane is due primarily to electrophoretic mobility. At certain pH values, the net electrical charge on a given bacterial cell is zero. When the cells are suspended in a fluid at this pH, they do not exhibit electrophoretic mobility; the pH at which this occurs is known as the isoelectric point. Increasing the pH on either side of the isoelectric point generally results in increased velocities of migrating cells.

Viable and non-viable spores of *Bacillus subtilis* var. *niger* were used to evaluate the ability of electrophoresis to differentiate between living and dead cells. Non-viable spore preparations were made by exposing viable spores to one of the following: ethylene oxide, pressurized steam, dry heat, chlorine, and formalin. Migration velocities of both living and dead spores were observed for varying pH levels above and below the isoelectric point, which was 2.5.

No reproducible differences between the viable and non-viable spores were indicated. On the basis of this study, it was concluded that the electrophoresis of living- and dead-cell preparations did not show differences of sufficient magnitude to provide a useful means for differentiating between viable and non-viable cells. A more sensitive means for measuring electrophoretic mobility as a function of pH would require refinement of the apparatus and techniques. However, based upon these experiments, it appears doubtful that even refinement will produce a sufficient magnitude of difference between living and dead cells to make electrophoresis a practical method for detecting small numbers of viable cells.

c. Autoradiography of cells exposed to a radioactive metabolite. A third method examined for its ability to detect viable cells was autoradiography. A cell which is metabolizing, but not necessarily reproducing, can be induced to "take up" metabolites labelled with certain radioisotopes, and thereby the cell becomes radioactive. By using autoradiography techniques, labelled cells can be induced to identify themselves by exposure to a photographic plate (Refs. 18 and 25). Autoradiography depends upon placing the labelled organism into intimate contact with a photographic emulsion that is sensitive to weak beta radiation. The emitted beta rays strike silver crystals in the emulsion, producing opaque "squiggles" at the site. The pattern of the opaque silver after development should be similar to the shape of the microorganism. When the sample is stained, the silver grain will be observed within the cell's boundary when examined microscopically.

To test autoradiography as a feasible method for detecting small numbers of viable microorganisms, viable and non-viable spores of *Bacillus subtilis* var. *niger* were used in individual tests. Tritiated (radioactive) thymidine, at a level of 0.25 μC, was added to separate suspensions of viable and non-viable cells in populations of 10^6 cells/ml of buffer-containing calcium chloride, dipicolinic acid, and tris. Similar cellular preparations were made without radioactive thymidine. The prepared samples were incubated for varying periods of time from 10 min to 6 hr. At the end of the incubation period, the suspension was centrifuged, and the resulting pellet of cells was resuspended in 5-ml quantities of sterile distilled water and again centrifuged. The pellet was then applied to a glass slide and spread in a thin film at the center of the slide. Following drying, the spores were fixed to the slides by dipping into a 1% solution of cellodion (ether:alcohol, 1:1) and thoroughly drying. A small drop of 1:1 egg-albumin–glycerine solution was then spread over the surface containing the spores. In the dark room, the dry slides were dipped in a photographic emulsion of Eastman Kodak Company's NTB-2 and NTB-3, drained, and allowed to dry in a vertical position. The thickness of the emulsion was determined to be approximately 4 μ. The slides were stored in a light proof container for 2 to 5 wk. After storage, the slides were developed with Eastman Kodak Developer DK 72.

Microscopic examination of the developed slides showed that, except for two cases, there was no evidence of thymidine on the autoradiographs. The two exceptions were slides that were in storage for the maximum time and, thus, the maximum time in contact with the emulsion. However, even in these cases the results were very weakly positive, and it was difficult to observe more than an occasional clump of several spores.

Based upon this study, it is concluded that autoradiography has very little potential as a technique for sterility testing of spacecraft components. In addition to the apparent low "take up" of radioactive thymidine by the spores, the presence of particles from a pulverized solid would create additional problems, because these particles would prevent the required intimate contact of the radioactive cell with the photographic emulsion. There is also a very real probability of artifacts resulting from particles binding the radioactive material. There appear to be too many disadvantages to this technique for use as a practical means of detecting microorganisms, especially spores.

d. Measurement of electron spin resonance. The fourth method examined as a possible means for detecting viable cells was the measurement of electron spin resonance

(ESR). The ESR phenomenon is the result of the absorption of energy by unpaired electrons under the influence of a magnetic field. ESR is concerned with the magnetic moments of unpaired electrons. Electrons have a spinning motion, and, since they are electrically charged, a magnetic field is created. When a spinning activated electron is subjected to a magnetic field at right angles to the main magnetic field which is pulsating, the frequency of the pulse is the same as the frequency of precession. Interaction will then occur which changes the orientation of the electrons, and energy is absorbed. The absorbed energy can be detected by the use of an ESR spectrometer.

Unpaired electrons appear to be essential components of living organisms by virtue of their necessary role in biological oxidation–reduction mechanisms (Refs. 20 and 24). The first evidence that unpaired electrons were associated with free radicals of metabolic origin was reported for lyophilized samples of microorganisms and tissues of plants and animals. Extremely sensitive instrumentation is required for the detection of ESR, since the free-radical concentration in most microbial systems is very low. The major sources of unpaired electrons in living systems include flavins, cytochromes, hemoglobin, myoglobin, certain photosynthesis enzymes, and perhaps desoxyribonucleic acid following excitation (Ref. 28).

In exploring the use of ESR as a feasible means for detecting viable cells, spores of *Bacillus subtilis* var. *niger* were placed in each of several quartz glass tubes. The spore population per tube was approximately 2.5×10^{1}. In a similar manner, yeast cells were placed in quartz tubes at a concentration of approximately 5×10^{9} cells/tube. The yeast cells and bacterial spores in a number of the tubes were treated with germicidal agents such as ethylene oxide, ethylene imine, and peracetic acid (40%).

The results of ESR detection were that both the treated and untreated bacterial spores gave a very strong signal (in excess of 1 mv). The yeast gave a weak signal that was the same for treated and untreated cells. Gentle heating of the tubes produced a darkening of the contents and an increase in signal; however, the same effect of heating and charring occurs with sugar.

If it were possible to use ESR as a reliable means for detecting viable cells in solid materials, it would offer the possible advantage of non-destructive testing. Unfortunately, these studies indicate that it is not possible to reliably distinguish between living and dead cells, or even between charred material and living material. The amplitude of the ESR signal in rapidly metabolizing and dead cells is not sufficient to be useful, even as a "screening" method.

In addition to the problems for biological materials, there are many solids of non-biological origin which produce strong, and consequently interfering, ESR signals. Therefore, it must be concluded that ESR spectroscopy is not a useful method for the detection of microorganisms, living or dead, in solid materials.

e. Culturing or growing of cells. The last method investigated was the culturing (growth) technique commonly used for sterility tests. Growth techniques have been the most reliable means of demonstrating the viability of microorganisms. The methods basically depend upon the ability or inability of an organism to reproduce when placed under assumed optimal conditions for growth and reproduction. Microorganisms which can reproduce make their presence discernible by increasing in numbers of new cells. This type of sterility test, which measures only the ability of the microorganism to reproduce and not its ability to metabolize independently of reproduction, may not be sufficient for evaluating spacecraft sterility. It is known that certain microorganisms in pharmaceutical and food products do continue metabolic processes and produce toxic by-products without increasing their numbers. It is a very real possibility that viable cells may be present in materials being tested, but are unable to reproduce and increase their numbers because of a number of factors (e.g., lack of proper nutrients, improper incubation conditions, presence of toxic materials, and sub-lethal cellular injury). Proper conditions can be created for specific microorganisms, but it is difficult to establish adequate conditions for the large variety of microorganisms that may be present. In general sterility testing, the latter condition usually applies.

For this study, growth conditions were investigated, and proper requirements were established for the organisms used in the investigation. These organisms were: (1) *Bacillus subtilis* var. *niger*, an aerobic spore-forming bacterium; (2) *Clostridium sporogenes*, an anaerobic spore-forming bacterium; and (3) *Ulocladium* sp., a fungus. Several commercially prepared nutrient growth media were screened, and it was found that satisfactory recoveries were obtained on trypticase soy medium for *Bacillus subtilis* var. *niger*, thioglycollate medium for *Cl. sporogenes*, and Sabouraud's dextrose medium for *Ulocladium* sp. Both room temperature (22 to 23°C) and 37°C were used as incubation temperatures.

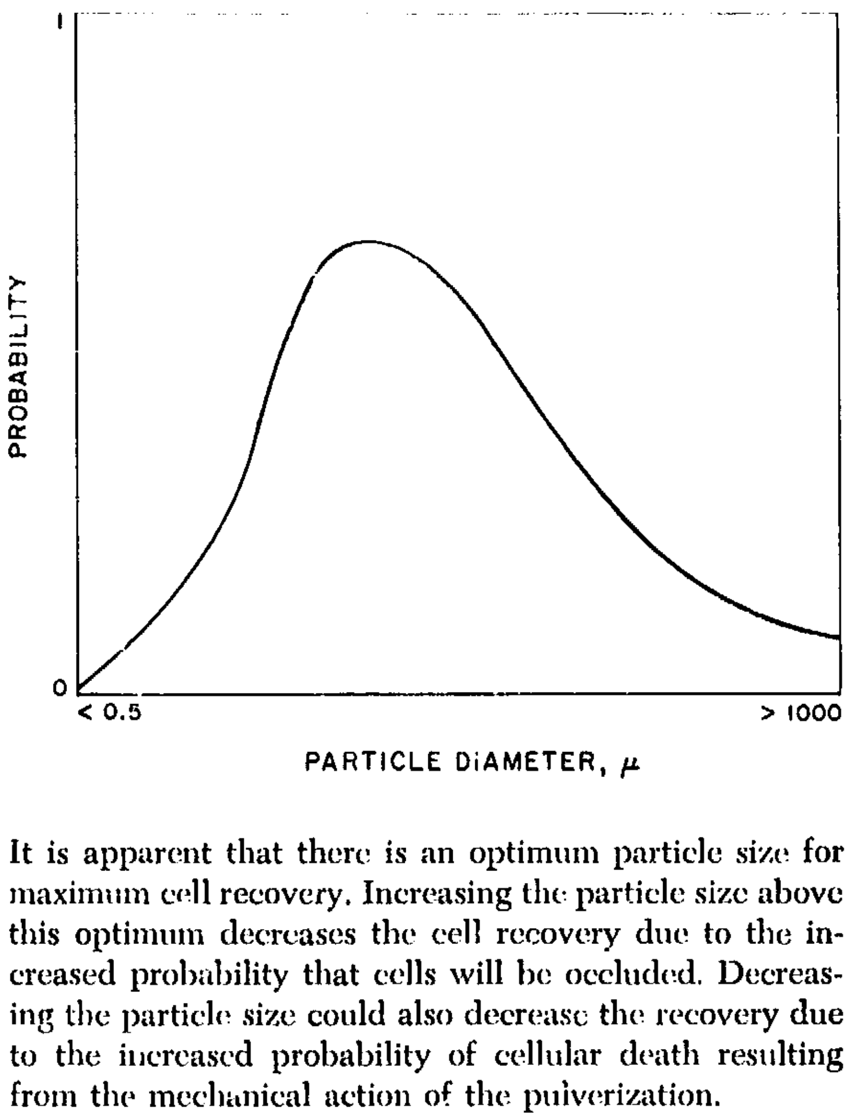
f. Results of Phase 1. The results of the study on the detection of viable microbial cells showed that the non-culturing techniques do not reliably distinguish between viable and non-viable cells. The culturing technique still appears to be the best available method to detect viability, even though many problems are encountered in the use of this technique for general sterility testing of solid materials. One of the greatest problems in the microbiological analysis of the interior of solids is the freeing of the cell from its encasing material so that detection of the cell can be optimized. This problem was considered in the next phase of the investigation.

2. Methods for Freeing Viable Microorganisms from Solid Materials

The primary objective of the second phase of the investigation was to develop a method for reliably freeing and enumerating viable microorganisms from the interior of solid materials with varying physical properties. As previously stated, the culture technique appears to be the best available means to differentiate between viable and non-viable microbial cells; therefore, this technique was selected as the detection method to be used for the remainder of the investigation. Detection of viable microorganisms by culturing requires that organisms increase their numbers by undergoing cellular division. For a cell to grow and divide, it must be in intimate contact with essential nutrients from which energy and new cellular material is derived. It is obvious that the cell must be free of substances which would interfere with its metabolic processes. Therefore, to adequately test for the presence of viable cells within solid materials, the cell must be freed of its encasement within the solid.

Basically, there are three methods that may possibly free microorganisms from the interior of solid materials: slicing, pulverizing, and dissolving. Dissolution of materials presents two major complex problem areas: (1) the extreme variability in solubility of materials in various solvents, and (2) the toxicity of suitable solubilizing agents to the viable microorganisms that might be present. Thus, it is apparent that dissolution is not a practical technique for sterility testing of spacecraft components, although it may have application in certain specific instances.

The techniques of slicing and pulverizing are, by no means, free of problems. Perhaps the greatest problem is establishing optimum particle size or slice thickness to give the maximum number of viable cells the opportunity to grow and reproduce under culturing conditions. A purely theoretical curve is shown below to illustrate the possible effects of particle size on the probability of recovery of viable cells using a pulverizing technique on solid materials.



It is apparent that there is an optimum particle size for maximum cell recovery. Increasing the particle size above this optimum decreases the cell recovery due to the increased probability that cells will be occluded. Decreasing the particle size could also decrease the recovery due to the increased probability of cellular death resulting from the mechanical action of the pulverization.

The following four types of equipment were used to comminute the solid materials used in the investigation: a ball mill, a blender, a drill, and a mortar and pestle. Four different types of pulverizing action are imposed on the solids by these methods. The ball mill provides a compressive impulse upon the solid, which tends to fracture during compression. The drill basically affords a cutting action on the solid, and the mortar and pestle provides a slow shearing force on solid particles. The pulverizing action of the blender is primarily the result of particles fracturing during impact, but a certain amount of shearing and cutting action may also be present.

The following materials were used in the exploratory investigations on pulverization and the recovery of viable microorganisms: (1) Paraplast (Brunswick Corporation, St. Louis, Mo.), (2) Parlodion (Mallin Krodt Chemical Company, St. Louis, Mo.), (3) Plaster of Paris, and (4)

Eccocoat IC2 (Emerson and Cuming, Inc., Canton, Mass.). These materials, representing a variety of physical characteristics, were selected in an attempt to simulate the properties of spacecraft materials that might harbor microbial contaminants. Specimens for examination were prepared by inoculating the liquid suspension of each material with spores of *Bacillus subtilis* var. *niger* and mixing thoroughly to obtain, as nearly as possible, a homogeneous mix. The suspensions were then poured into small aluminum molds for curing. The materials used in the preparation of the specimens were demonstrated not to suppress the growth of the test organism by the usual bacteriostasis tests, i.e., exposing the test organisms to the material for various periods of time and then culturing the organisms. Aseptic and/or sterile techniques were used throughout the investigations to prevent or minimize extraneous microbial contamination.

The inoculated specimens were pulverized using the procedures previously indicated. In the drilling technique, the cutting action was provided by a ³⁄₃₂-in.-D spherical-type carbide bit affixed to a flexible shaft on a high-speed drill motor (Dremel Manufacturing Company, Racine, Wisc.). The ball mill was a No. 5000 mixer mill (Spec Industries, Inc., Scotch Plains, N. J.), in which the pulverizing chamber consisted of a hardened alloy steel vial (Spex No. 3124) containing a stainless-steel ball. The mortar and pestle were the porcelain-type; the mortar was a 1-qt Waring blender with a Pyrex container (Waring Products Company, Winsted, Conn.).

The particles resulting from each pulverizing technique were suspended or dissolved in appropriate non-toxic fluids. An amount of material calculated to contain approximately 10^6 spores was used for each suspension prepared.

Trypticase soy agar (Baltimore Biological Laboratories, Baltimore, Md.) plates inoculated with aliquots of the fluids were then prepared and incubated at 37°C for 7 days. Colony counts were made on the incubated plates, and the counts were used as the basis for evaluating the microbial recovery capabilities of each of the pulverizing techniques. Suitable control tests were conducted throughout the investigations to determine the viability of the spore inoculum and any bacteriostatic effects due to materials or procedures used in the experiments.

The results of particle-size measurements are shown in Table 6, where maximum particle diameters are given for the four solid materials pulverized by the four different techniques. Although a particle-size distribution was not reported, it would appear that the drill technique produced smaller particle sizes than the other methods. However, it is also apparent that particles of the sizes reported here could still harbor large numbers of bacteria, considering the average diameter of a bacterium to be approximately 1 μ.

Table 6. Maximum particle sizes obtained by various pulverizing techniques

Pulverizing equipment	Grinding time, min	Maximum particle diameter, μ			
		Paraplast	Parlodion	Plaster of Paris	Eccocoat IC2
Drill	0.5	150	40	100	40
Ball mill	2.0	150	1000	100	1000
Mortar and pestle	1.0	150	1000	100	1000
Blender	1.0	150	400	100	400

The recovery of viable spores from the pulverized materials is shown in Table 7. The data indicate that the drill technique gave somewhat higher and more consistent recovery than did the other methods, although in certain isolated instances it would appear that one of the other techniques gave a better recovery. Recovery at best is only about 1%, and, in most cases, the data indicate less than a 1% recovery. In several instances, the recovery from replicates of the same technique shows marked differences in the number of colonies counted. The exact reasons for these variations are not known. However, it is possible to postulate several reasons, including nonuniform distribution of the spores inoculated into the materials and variations of the individual performing the techniques. For example, differences in pressures applied when using the drill or the mortar and pestle can result in differences in the size of the particles produced, which would result in variations in numbers of cells freed from encapsulation. The variations in particle size can occur even though the maximum particle size may be essentially the same for each material.

An analysis of variance was performed on the data to determine which of the variables of the procedures exerted the greatest influence on particle size and recovery. The results of this analysis indicate that, under the conditions of this study, the type of material had more effect upon particle size and recovery than did the other variables such as pulverizing technique, specimen treatment, and replication. It should be noted, however, that the effect of the material on recovery is much less noticeable for the softer materials such as Paraplast and Plaster of

Table 7. Number of *Bacillus subtilis* var. *niger* colonies recovered from pulverized solids

Material	Specimen treatment	Replicate	Pulverization technique							
			Drill[a]		Ball mill		Mortar and pestle		Blender	
			Run 1	Run 2	Run 1	Run 2	Run 1	Run 2	Run 1	Run 2
Paraplast	Pulverized and suspended	a	5,750	6,680	—	3,600	6,990	6,280	1,880	1,950
		b	6,080	2,120	300	350	2	5,750	2,150	1,900
	Pulverized and dissolved	a	2,750	2,840	2,430	2,490	5,600	4,850	8,900	8,860
		b	6,080	6,080	4,020	3,850	1,150	7,800	8,200	7,600
	Dissolved; no pulverizing	a	8,300	8,160	—	—	—	—	—	—
		b	3,140	6,800	—	—	—	—	—	—
Parlodion	Pulverized and suspended	a	120[b]	120	2	1	220	200	56	41
		b	180[b]	150	0	3	4	4	40	28
	Pulverized and dissolved	a	60[b]	20[b]	1,850	1,820	0	0	35	16
		b	140[b]	120[b]	60	40	4	31	42	0
	Dissolved; no pulverizing	a	7,190	8,900	—	—	—	—	—	—
		b	9,790	10,000	—	—	—	—	—	—
Plaster of Paris	Pulverized and suspended	a	10,750	10,100	4,470	6,280	6,600	5,490	5,350	6,100
		b	10,070	8,200	120	150	4,800	4,500	4,000	4,140
	Pulverized and dissolved	a	10,600	10,900	7,500	7,850	7,900	6,250	9,430	9,750
		b	780	740	8,900	6,800	8,420	10,000	8,500	9,400
	Dissolved; no pulverizing	a	7,000	11,000	—	—	—	—	—	—
		b	6,800	6,100	—	—	—	—	—	—
Eccocoat IC2	Pulverized and suspended	a	1,580	1,580	0	1	1,220	1,020	5	5
		b	150	260	5	7	46	61	3	5
	Pulverized and dissolved	a	250	210	1	0	1	1	22	10
		b	1,240	1,540	0	0	6	9	1	2
	Dissolved; no pulverizing	a	7,850	7,200	—	—	—	—	—	—
		b	7,000	8,000	—	—	—	—	—	—

[a]Replicate agar plate.
[b]The inoculum in the solid actually tested for microorganisms was 10^5 spores; all other specimens contained approximately 10^6 spores.

Paris, and recovery by the drill technique appears to be only slightly better from these materials.

3. Conclusions

From the results of these studies on detection methods for viable microbial cells, it is concluded that the most reliable means of differentiating between viable and non-viable microorganisms is the use of the culturing (growth) technique. The physical methods (staining, electrophoresis, autoradiography, and electron spin resonance) did not reliably differentiate between viable and non-viable cells as required in sterility testing. In addition, they present numerous problems that make them impractical for general use.

Of the four techniques studied for comminuting solid materials, it must be concluded that the drill technique gives the best recovery of viable cells. However, the drill technique, as used in this investigation, still leaves much to be desired since maximum recovery is only at the 1% level. The reliability and applicability of this technique to a broader range of solid materials are to be evaluated.

The data collected in this investigation indicate that there are several problem areas which must be resolved; therefore, continuing investigations to improve techniques for the recovery of viable cells from solids are required if practical and reliable sterility test procedures are to be developed to support certification of spacecraft sterility.

References

1. Michaelson, G., and Vesley, D., "Industrial White Rooms vs. Hospital Operating Rooms," *Air Engineering*, Vol. 5, pp. 24–29, 1963.

2. *Microbial Contamination Obtained on Surfaces Exposed to Room Air or Touched by the Human Hand*, Protection Branch Report of Test No. 1-64, Physical Defense Division, Fort Detrick, Frederick, Maryland, July 22, 1963.

3. *The Level of Microbial Contamination In a Clean Room During a One Year Period*, Protection Branch Report of Test No. 11-65, Physical Defense Division, Fort Detrick, Frederick, Maryland, December 4, 1964.

4. Bateman, J. B., McCaffrey, P. A., O'Connor, R. J., and Monk, G. W., "Relative Humidity and the Killing of Bacteria. The Survival of Damp *Serratia marcescens* in Air," *Applied Microbiology*, Vol. 9, pp. 567–571, 1961.

5. Dunklin, E. W., and Puck, T. T., "The Lethal Effect of Relative Humidity on Airborne Bacteria," *Journal of Experimental Medicine*, Vol. 87, pp. 87–101, 1948.

6. Kethley, T. W., Fincher, E. L., and Cown, W. B., "The Effect of Sampling Method Upon the Apparent Response of Airborne Bacteria to Temperature and Relative Humidity," *Journal of Infectious Diseases*, Vol. 100, pp. 97–102, 1957.

7. Lidwell, O. M., and Lowbury, E. J., "The Survival of Bacteria in Dust. II. The Effect of Atmospheric Humidity on the Survival of Bacteria in Dust," *Journal of Hygiene*, Vol. 48, pp. 21–27, 1950.

8. McDade, J. J., and Hall, L. B., "An Experimental Method to Measure the Influence of Environmental Factors on the Viability and the Pathogenicity of *Staphylococcus aureus* in the Environment," *American Journal of Hygiene*, Vol. 77, pp. 98–108, 1963.

9. McDade, J. J., and Hall, L. B., "Survival of *Staphylococcus aureus* in the Environment. I. Exposure on Surfaces," *American Journal of Hygiene*, Vol. 78, pp. 330–337, 1963.

10. McDade, J. J., and Hall, L. B., "Effect of the Environment on Virulence," National Conference on Institutionally Acquired Infections, *U. S. Public Health Monograph No. 1188*, pp. 91–97, 1963.

11. McDade, J. J., and Hall, L. B., "Survival of *Staphylococcus aureus* in the Environment. II. Effect of Elevated Temperature," *American Journal of Hygiene*, Vol. 80, pp. 184–191, 1964.

12. McDade, J. J., and Hall, L. B., "Survival of Gram-Negative Bacteria in the Environment. I. Effect of Relative Humidity on Surface-Exposed Organisms," *American Journal of Hygiene*, Vol. 80, pp. 192–204, 1964.

13. McDade, J. J., "Sources of Microbiological Contamination Within Clean Rooms," *Proceedings of the American Association for Contamination Control, Session XI, Paper 3*, 1964.

14. Hoffman, R. K., *A Technique for the Investigation of Bacterial Contamination Inside Electronic Components*, Protection Branch Report of Test No. 7-60, Physical Defense Division, Fort Detrick, Frederick, Maryland, December 19, 1962.

References (Cont'd)

15. Meisel, M. N., Medvedeva, G. A., and Alekseeva, V. M., "Discrimination of Living, Injured, and Dead Microorganisms," *Mikrobiologiya*, Vol. 30, p. 855, 1961.

16. "Fluorescence Microscopy with Fluorochromes—Recipes and Tables," *Reichert Optische Werke AG*, Vienna, Austria, 1963.

17. Bangham, A., Heard, D., Flemans, R., and Seaman, G., "An Apparatus for Microelectrophoresis of Small Particles," *Nature*, Vol. 182, pp. 642–644, 1958.

18. Baserga, R., and Nemeroff, K., "Factors Which Affect Efficiency of Autoradiography with Treated Thymidine," *Stain Technology*, Vol. 34, pp. 21–26, 1962.

19. Brinton, C., and Lauffer, M., "The Electrophoresis of Viruses, Bacteria, and Cells and the Microscope Method of Electrophoresis," *Electrophoresis*, Bier, M., Ed., Academic Press, Inc., New York, 1959.

20. Commoner, B., Townsend, J., and Pake, G. E., "Free Radicals in Biological Materials," *Nature*, Vol. 174, p. 689, 1954.

21. Douglas, H. W., "Electrophoresis Studies on Bacteria. Part 5. Interpretation of the Effects of pH and Ionic Strength on the Surface Charge Borne by *B. subtilis* spores, with Some Observations on Other Organisms," *Transactions of the Faraday Society*, Vol. 55, pp. 850–856, 1959.

22. Gurr, E., *Staining: Practical and Theoretical*, Williams and Wilkins Co., Baltimore, Maryland, 1962.

23. Harris, J. O., and Kline, R. M., "Electrophoresis of Motile Bacteria," *Journal of Bacteriology*, Vol. 72, p. 530, 1956.

24. Heckley, R. J., Dimmick, R. L., and Windle, J. J., "Free Radical Formation and Survival of Lyophilized Microorganisms," *Journal of Bacteriology*, Vol. 85, p. 961, 1963.

25. Joftes, D., "Radioautography, Principles and Procedures," *Journal of Nuclear Medicine*, Vol. 4, pp. 143–154, 1963.

26. Kuhn, W. E., Lamprey, H., and Sheer, C., *Ultrafine Particles*, John Wiley & Sons, Inc., New York, 1963.

27. Poro, T. J., et al., "Fluorescence and Absorption Spectra of Biological Dyes," *Stain Technology*, Vol. 38, p. 37, 1963.

28. Szent-Gyorgyi, A., *Introduction to Submolecular Biology*, Academic Press, Inc., New York, 1960.

GUIDANCE AND CONTROL DIVISION

IV. Spacecraft Secondary Power

A. Power Sources

S. S. Luebbers

1. Thermionic Converter Development

a. Introduction. The technical effort presented in this summary covers the first phase of the development of a high-performance thermionic converter, and describes the work performed under JPL Contract 950699 with Electro-Optical Systems, Pasadena, California.

The performance goal of the program was to produce a thermionic converter capable of delivering 20 w/cm^2 electrical power. This power was to be obtained at an output voltage of 0.7 v, at a true emitter temperature of 1700°C. In attempting to reach these performance goals, seven different iterations of a basic EOS thermionic converter were performed. Some of the iterations led to improved performance, while others resulted in lower performance than anticipated. In this summary, the results of the seven iterations are discussed along with the technical aspects involved.

Before undertaking the task of developing a high-performance converter, EOS had been able to produce, by what shall be termed basic converter production techniques, a converter capable of delivering 12 to 13 w/cm^2. This power density represents a little more than 30 w of electrical power per converter; however, this figure is below that which has been demonstrated to be achievable. The following four converter parameters were studied and are discussed in this summary: (1) radiator heat rejection and heat load, (2) electrode spacing, (3) collector work function, and (4) emitter work function.

b. Converter S/N 1 fabrication and performance. Testing of previous EOS converters had indicated that the radiator design was not adequate to meet the demands of high-current operation. To reduce the heat load which the radiator must reject, the emitter area was reduced approximately 25%. This reduced the heat transferred to the collector and consequently to the radiator and allowed the utilization of available radiator parts. Due to the reduced heat load, the collector temperature was optimized.

In addition to the reduction of heat load, the converter electrode spacing was reduced from 2 mils to 1.5 mils. Experimental results from a previous contract with EOS indicated reduced spacing would lead to higher performance. Electrode spacing versus power output is illustrated in Fig. 1. (Note the earlier power outputs.)

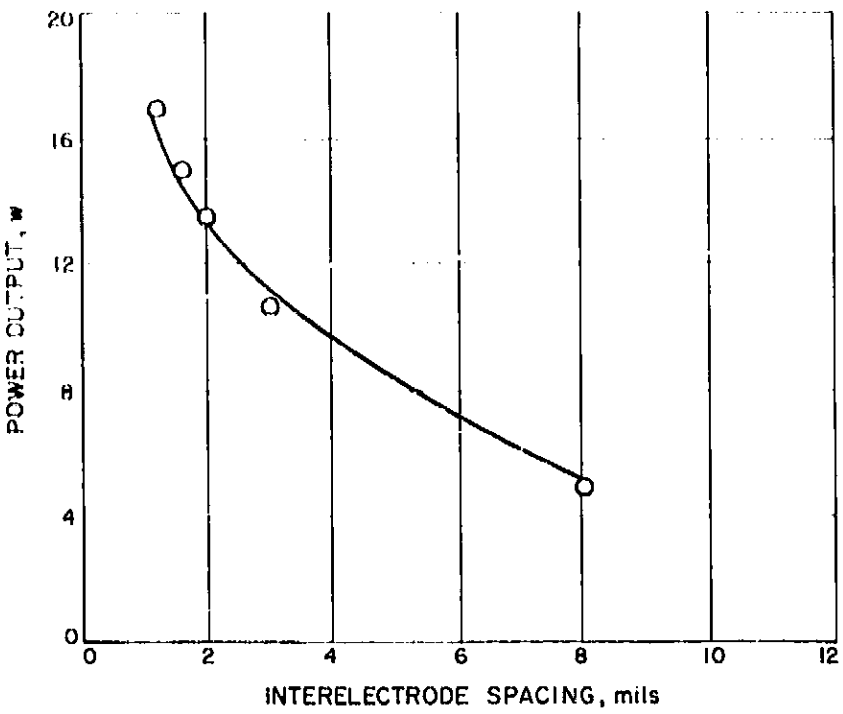


Fig. 1. Power output versus spacing

The reduction in spacing was achieved by changing the fabrication techniques. Fig. 2 is a drawing of the EOS converter. To obtain reduced spacing, the final assembly procedure was altered from the production converter techniques. Rather than seal brazing and then welding the emitter to the envelope, the process was reversed. It was felt that the emitter envelope would be prestressed in this manner, and smaller interelectrode spacing would result.

Initial testing of S/N 1 yielded a power density of 16 watts/cm² (Table 1) under the prescribed operating conditions. The power density had been significantly improved as anticipated. The volt-ampere curve for S/N 1 is presented in Fig. 3.

After 40 hr of operation, the vacuum system was vented to atmosphere. Upon resumption of testing, the converter showed a degradation in power output of 53%. This degradation is believed to have occurred due to a marginal seal braze. Upon exposing the converter to atmosphere, a small leak resulted in the contamination of the electrode surfaces. It is felt this leak resulted from the prestressed condition of the envelope.

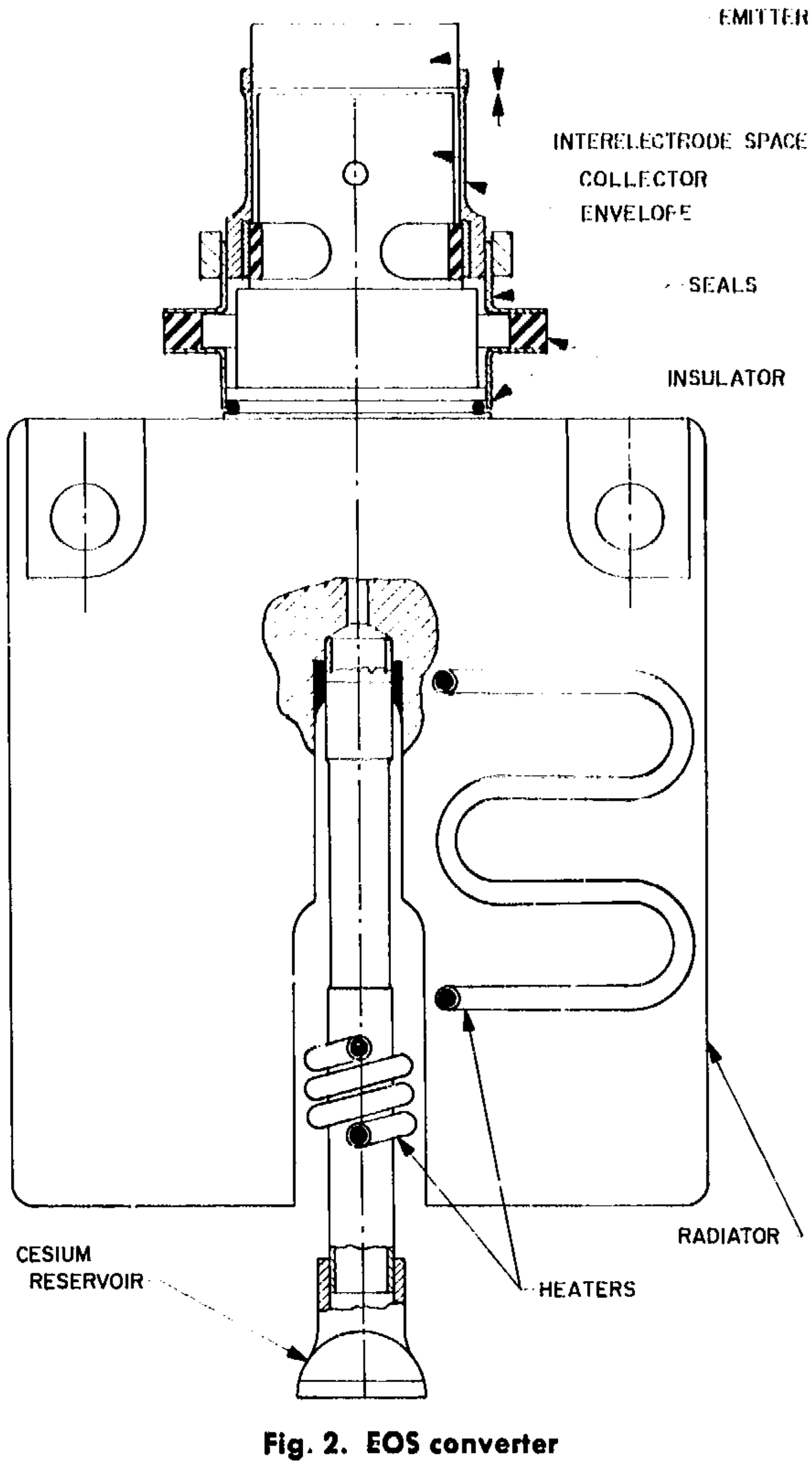


Fig. 2. EOS converter

c. Converter S/N 2 fabrication and performance. The marginal seal braze experienced with S/N 1 suggested that the original assembly procedure be readopted. S/N 2 was fabricated according to the established assembly procedure with no attempt to reduce spacing.

The performance of S/N 2 (Table 1) did not show significant improvement over that of the earlier production converters. The reduced heat load (smaller emitter) did, however, allow the optimization of the collector temperature at high currents.

Presently S/N 2 is undergoing life-testing at JPL. After operation for over 1500 hr, this converter has not experi-

Table 1. Summary of converter performances

Converter	Special features	Performance at 0.7 v
SN-1	Electrodes, Ta-Mo; long envelope seal brazed last to achieve closer spacing	16.0 w/cm^2 → 8.1 w/cm^2 at 40 hr; degradation due to small leak
SN-2	Electrodes, Ta-Mo; long envelope with original assembly procedure	12.5 w/cm^2; no degradation, more reliable assembly procedure, exhibited wider spacing
SN-3	Electrodes, Ta-Re; long envelope with original assembly procedure	15.2 w/cm^2 → 8.9 w/cm^2 at 40 hr; Re contaminated by Ta evaporation
SN-4	Electrodes, Ta-Re; shorter envelope assembly to achieve closer spacing	17.8 w/cm^2 → 5.4 w/cm^2 at 41 hr; degradation due to leak
SN-5	Electrodes: Re-Re; shorter envelope; 0.1 to 0.2-mil Re on emitter; close spacing	No initial performance data obtained; after degradation due to diffusion, output stabilized at 14.5 w/cm^2
SN-6	Electrodes, Re-Re; shorter envelope; 20-mil Re on emitter; close spacing	Performance at 0.7 v poor due to I-V curve shift; 18.8 w/cm^2 at 0.6 v output; performance stabilized
SN-7	Same as SN-6	16.4 w/cm^2 at 1700°C, 20.0 w/cm^2 at 1723°C, performance stabilized

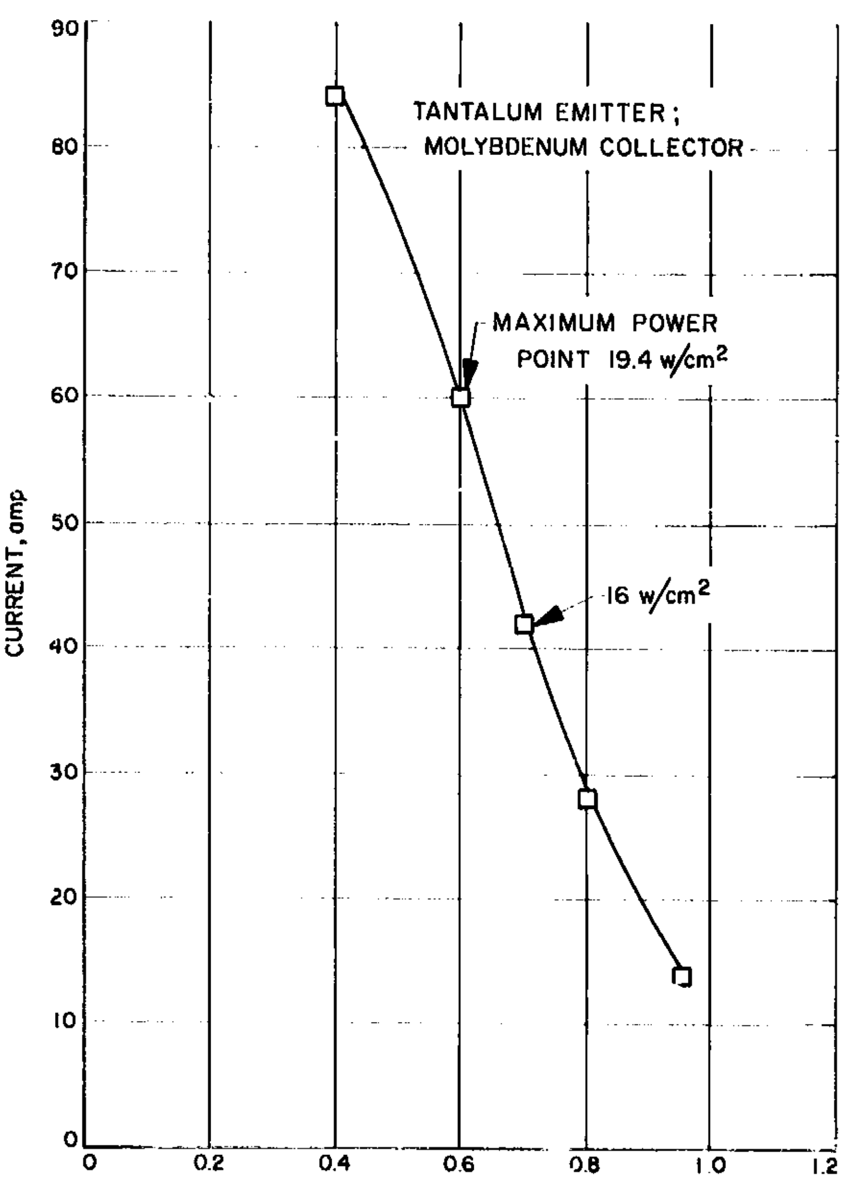


Fig. 3. Volt-ampere characteristics of converter SN-1

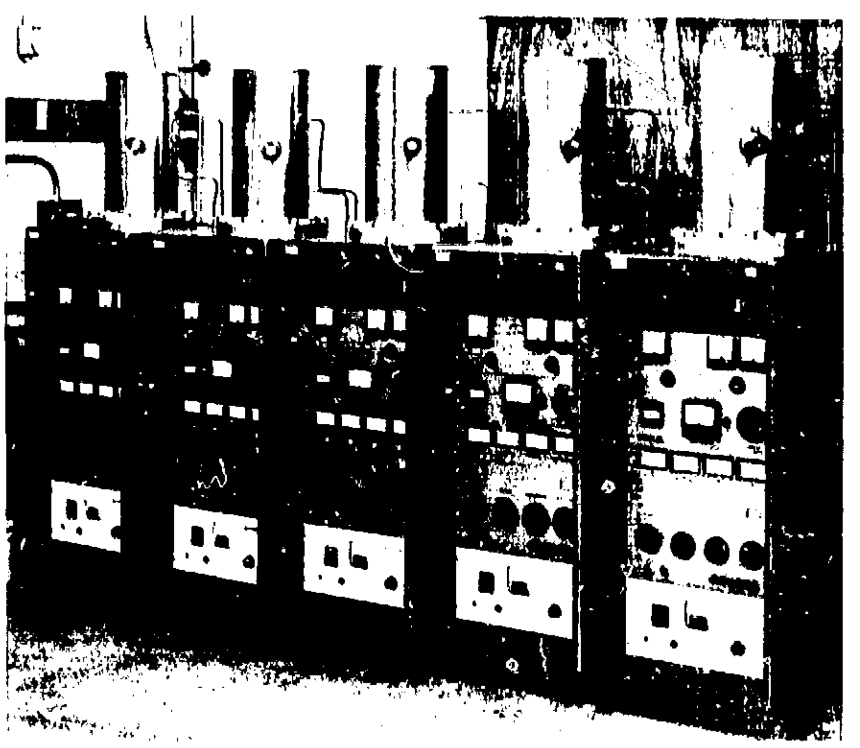


Fig. 4. Converter life test stations

enced any degradation. Converter life test stations are illustrated in Fig. 4.

d. Converter S/N 3 fabrication and performance. Converter S/N 3 was fabricated to assess the performance of rhenium as a lower collector work function material. With the exception of a thin rhenium coating which was evaporated onto the collector, the fabrication of S/N 3 was identical to S/N 2.

The effect of the work function ϕ on the output voltage of a thermionic converter may be expressed by

$$V_{out} = \phi_{emitter} - \phi_{collector} - V_p$$

where $\phi_{emitter}$ is the emitter work function, $\phi_{collector}$ the collector work function, and V_p the plasma sustaining voltage. From the above equation, a reduced collector work function is seen to be desirable. Several material studies had indicated that utilization of rhenium in a thermionic diode will result in a reduction of the collector work function by 0.1 to 0.15 ev over tantalum and, hence, serves as a basic reason for applying rhenium to the collector of converter S/N 3.

The performance of converter S/N 3 (Table 1) is plotted in Fig. 5. The initial performance of S/N 3 was significantly improved over that of the previous converter. The rhenium collector did improve the performance; however, this improvement was only temporary. After 140 hr of operation, the performance degraded by approximately 50%.

It appears that the mechanism of degradation can be explained by a contamination of the rhenium collector with tantalum. The emitter material, tantalum, is heated

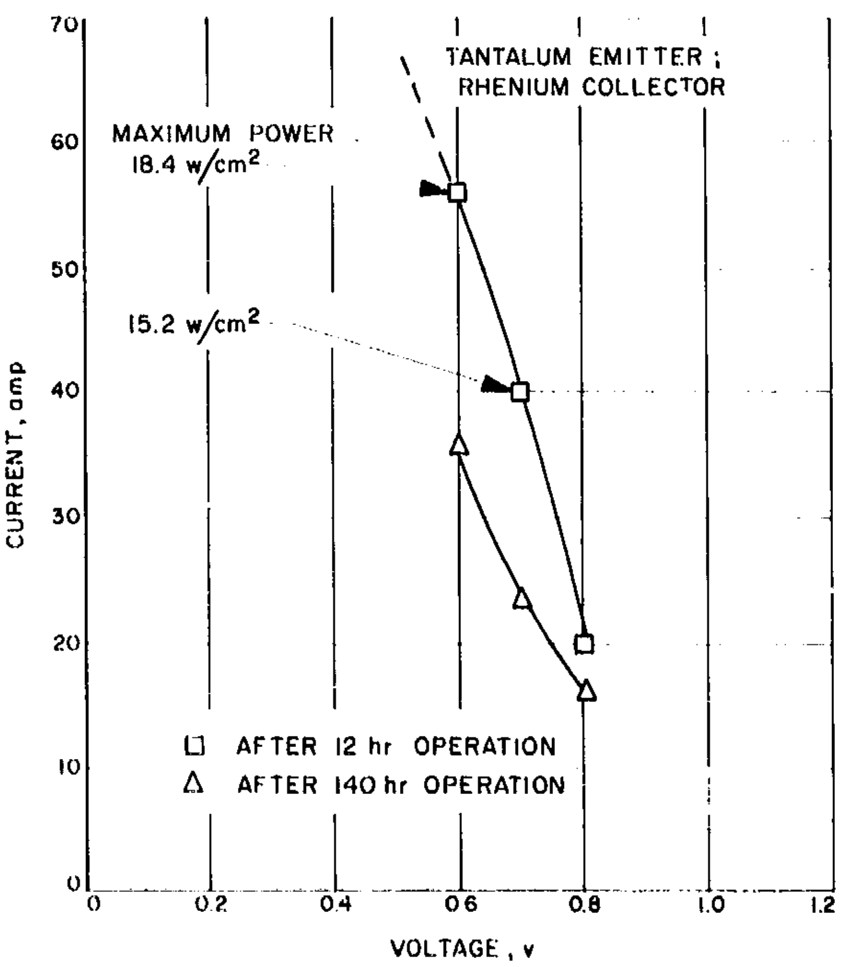


Fig. 5. Volt-ampere characteristics of converter SN-3 at 1700°C

to 1700°C. At this elevated temperature, tantalum evaporation cannot be avoided. The evaporated tantalum is transferred to the cooler collector and is condensed, where it serves to increase the work function of the collector.

e. Converter S/N 4 fabrication and performance. To benefit from the advantages gained by closer electrode spacing and a reduced collector work function, converter S/N 4 was fabricated similar to S/N 3 with the exception of a reduced envelope length. The shorter envelope allows less expansion between the emitter and collector surfaces as the converter is brought to temperature. In this manner reduced interelectrode spacing is obtained. It was also hypothesized that if the tantalum collector coating hypothesis was correct for converter S/N 3, degradation should occur on S/N 4 after extended operation.

The initial performance from converter S/N 4 (Table 1) was 17.8 watts/cm² under the prescribed operating conditions. After 1 hr of operation, testing was terminated to make some minor modifications within the vacuum system. Upon resumption of testing, it was observed that the converter performance had degraded approximately 50%. This behavior is characteristic of the degradation experienced with converter S/N 1, and indicated that a small leak had developed in the converter.

The initial performance of S/N 4 did support the expected performance level. The reduced collector work function and decreased spacing yielded the highest performance achieved by an EOS converter to that time.

f. Converter S/N 5 performance. To eliminate the degradation experienced with converter S/N 3, the emitter of S/N 5 was coated with a thin film (0.1 mil) of rhenium. The rhenium coating was applied to the emitter to reduce the transport of emitter material to the collector. Similar to S/N 4, the converter was fabricated with a shorter envelope assembly to reduce the interelectrode spacing.

Initial test results from converter S/N 5 (Table 1) appeared extremely promising; however, before all the parameters could be optimized, the performance fell to 14.5 w/cm². This performance is characteristic of a tantalum emitter and rhenium collector as noted in converter S/N 3. It is assumed that the initial coating of rhenium on the emitter surface diffused into the tantalum substrate. By this diffusion process either a tantalum or a tantalum-rhenium intermetallic surface was formed. This, in turn, resulted in lower emission performance.

g. Converters S/N 6 and S/N 7 performance. In order to prevent tantalum-rhenium diffusion or at least significantly reduce the diffusion rate, it was determined that a sheet of rhenium be applied to the emitters of the next iterations. Both S/N 6 and S/N 7 were fabricated with 20-mil rhenium sheet pressure bonded to the emitter surface. Other converter dimensions were not changed from the previous iteration.

Test results from converter S/N 6 (Table 1) indicated improved power performance at 0.6 v output, and lower performance at 0.7 v output then obtained from S/N 4. The volt-ampere characteristic had increased; however, at the same time it shifted toward the current axis. The shift of the curve toward the current axis reduced the performance at 0.7 v output; while, at 0.6 v output, the performance increased to 18.8 w/cm². The reason for this shift has not been clearly established.

The test results obtained from converter S/N 7 (Table 1) are presented in Fig. 6. As indicated in the figure, the finally achieved performance closely approached the goal of 20 w/cm² at 0.7 v. Converter S/N 7 employed a variation of all four parameters described in the introduction. The performance of this final converter demonstrated the importance of parametric control in thermionic converter development. By simply controlling several basic converter parameters, the program described herein has demonstrated an improvement of 25% in converter output performance.

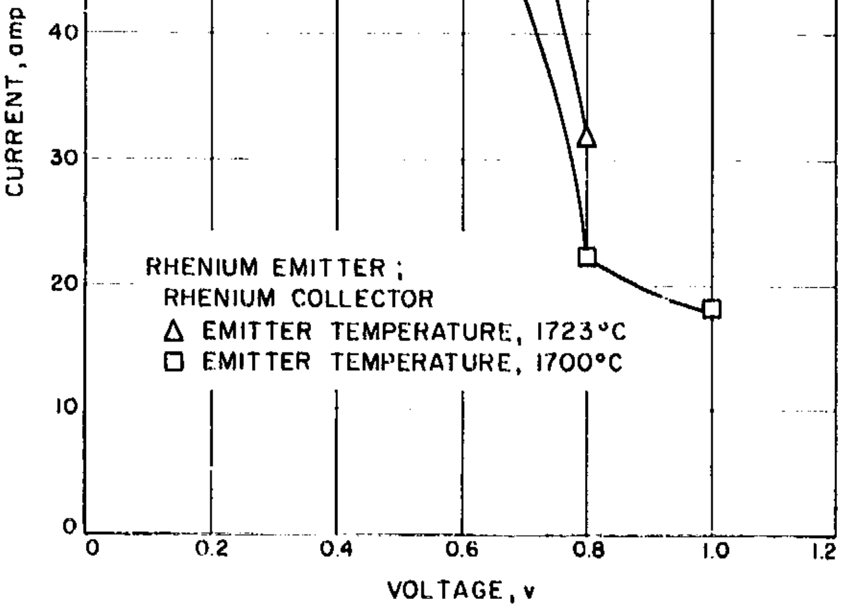


Fig. 6. Volt-ampere characteristics of converter SN-7

B. Electrical Conversion

R. A. Booth and A. Schloss

1. Solid State Switch

a. Summary. The purpose of the solid state switch development is to replace the existing motor-driven switch with its attendant problems, such as weight, size of external field, cost, reliability, and delivery. The described circuit is a hybrid approach (Fig. 7). In this switch, a transistor circuit handles the turn-on and turn-off transients and allows only 3 volts or less to appear across an associated pair of relay contacts during contact transfer. The relay contacts are closed for the steady state load and provide for high efficiency.

Relay contacts are rated for a combination of current and voltage. The contact cycle life is guaranteed for the specified combination. The open-circuit voltage determines the electric field strength across the contacts and this may be exceedingly high when the contacts are just breaking or just closing. This high field strength is one of the factors which tends to reduce cycle life. Current may be traded off against voltage, however the usual contact specifications do not reflect this. By holding the open circuit voltage to 3 volts or less during transfer, unusually high currents at reasonable cycle life can be handled by a small relay contact. After transfer, the open-circuit voltage may be increased as this is only limited by the contact insulation. Large relays capable of handling the required transient and steady state load reliably are not available.

At the present time, a hybrid solid state switch has been designed and a breadboard fabricated (refer to *SPS 37-30*, Vol. IV, p. 28), from which test results have been obtained at −10, 25, and 75°C.

The design was completed using only high reliability components, consistent with the present state-of-the-art. This design was also optimized for efficiency and reliability, and an effort was made to reduce the total number of components used.

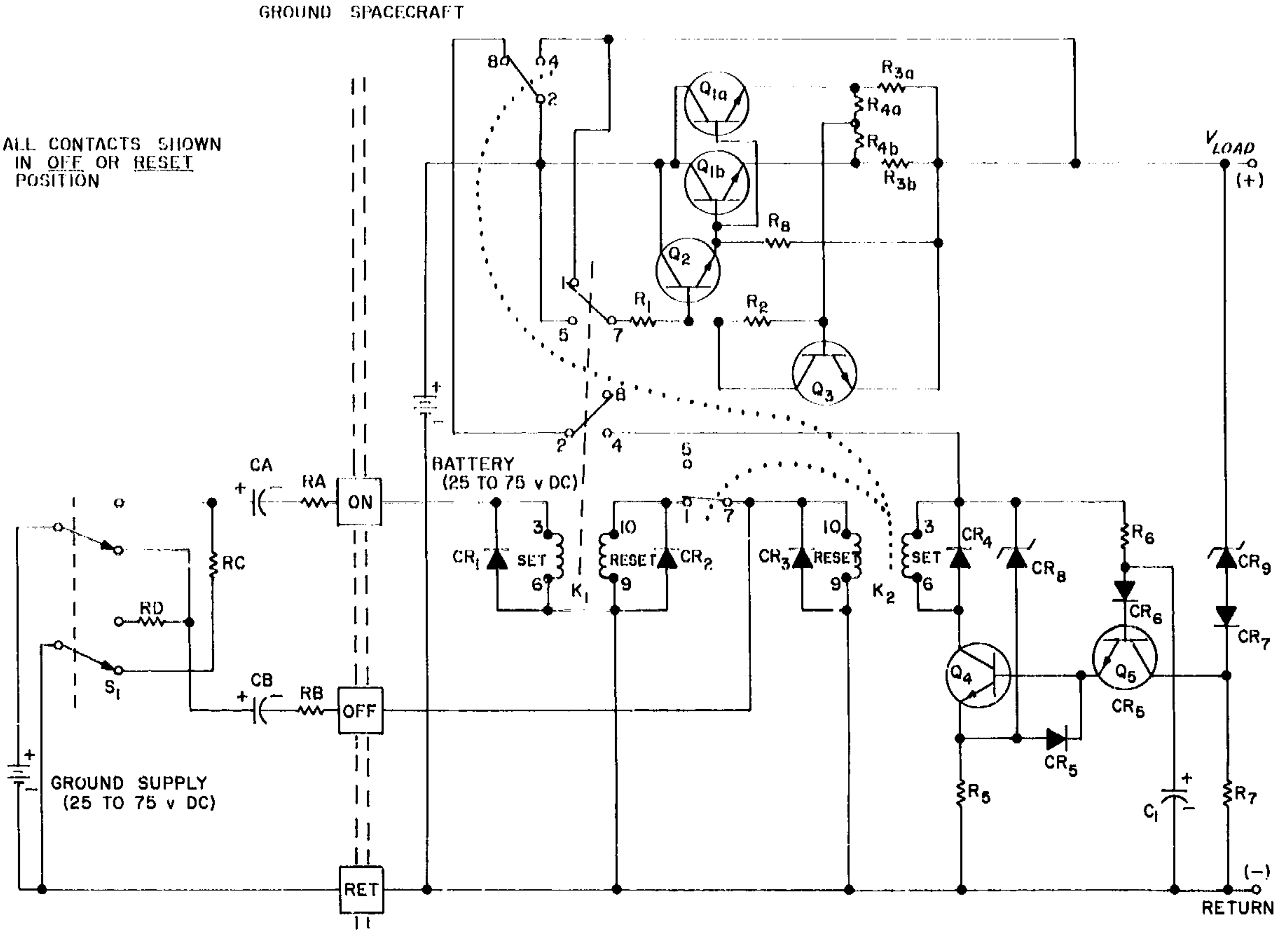


Fig. 7. Hybrid solid state switch schematic diagram

b. Operation. The operation of the hybrid solid state switch is as follows:

With all contacts as shown in the schematic (Fig. 7), the switch shall be considered as "off" or in the reset mode. In the ground support equipment (left of the interface lines), C_A and R_A provide a means to set K1. R_C provides a means of discharging C_A after its use. C_B and R_B are used to turn off (reset) both K2 and, after a slight delay, K1. R_D provides a means of discharging C_B after its use.

When K1 (on the spacecraft side of interface lines) is set by ground command as described above, contacts 7 to 5 close, and Q2 is driven-on by the spacecraft battery. The parallel combination of Q1a and Q1b constitutes a solid state switch which handles the turn-on and turn-off transients. R3a and R3b are balancing resistors and tend to compensate for the beta spread of Q1a and Q1b. Q2 provides base current drive to both Q1a and Q1b. R1 allows sufficient base current to drive Q2 fully on.

Q3 provides for current limiting during the time that the load capacitance is charging. The average load current is measured by sensing the voltage developed between the junction of R4a and R4b and the junction of R3a and R3b. When the load current exceeds a prescribed maximum value above the maximum nominal load, the base of Q3 is forward-biased sufficiently to drive Q3 into saturation, this starves the base of Q2 and, as a result, limits the collector currents of both Q1a and Q1b. Relay K1 switch contacts (terminals 2 to 4) are closed at the same time 7 and 5 are closed and provide a path from the S/C battery to the K2 "set" coil. Regardless of the state of K1, K2 will *not* "set" unless certain conditions are met; to wit, the voltage drop from the collectors of Q1a and Q1b to the junction of R3a and R3b (when K1 is set) must be 3.0 v DC or less. This is necessary so that relay K2 switch contacts (terminals 2, 4) will only start to close when the applied voltage across its contacts is 3 volts or less. When the relay K2 switch contacts (terminals 2, 4) *do* close, the potential as seen across the contacts is the summation of Q1a or Q1b saturation voltage and R3a or R3b voltage drop. This mechanization provides the desired contact life at high current levels (20 amp maximum).

The desired condition then that must be met before K2 sets, is that its contacts see 3 volts or less and this, regardless of whether the battery has any terminal voltage from 25 to 75 volts. The key to the logic that implements this is the difference in zener voltages of CR8 and CR9. Neglect for the moment the action of C1. The load may be highly capacitive and so the output (load) point may not reach its steady state value the instant K1 is made to set. At the instant of set, CR8 conducts and has the design voltage value across it (V_{CR8}). The load voltage rises until there is sufficient voltage to make CR9 conduct, at which time it will have its design voltage value across it (V_{CR9}). When $V_{CR9} + V_{CR7} + V_{Q5SAT} + V_{BEQ4} + V$ across terminals 2 to 4 of K2 equals V_{CR8}, Q4 conducts and K2 will set. In the above balance, V_{CR9} and V_{CR8} are chosen such that the voltage across the relay contacts will be 3 volts or less when Q4 conducts. It can be seen that the above balance may be obtained (withing design limits) for any applied battery voltage. C1 may prevent the setting of K2 until a later time if the R6, C1 time constant is greater than the current limited charge time of the load capacity. (C1 is provided to alleviate a reset problem.) When K2 sets, terminals 2 to 4 contact, providing a low ohmic bypass of the solid state transient switch; terminals 2 to 8 open which prevents continued battery drain through Q4 and finally (when C1 discharges) through Q5. At the same time, terminals 1 to 7 of K2 open which merely sets the proper condition for the reset sequence.

The turn-off sequence reverses the above, the object here being that terminals 2 to 4 shall open to no more than 3 volts. The 3 volt condition must remain until the contacts are fully open. The contact life is determined by "open to" voltage for this determines the electric field strength at the early opening stage. Once fully opened, only the contact insulation determines what the voltage may be.

Ground voltage is applied through C2 to the reset winding of K2. C2 is large enough to insure enough movement of total charge required to reset the relay. Since K1 is still set at the beginning of this sequence, base drive to Q2 is available. When K1 opens, the conditions will be as before and K1 contacts will see less than 3 volts. (Q1 saturation drop and R3 drop) K2 contacts 1 to 7 close as contacts 2 to 4 open. The bypass contacts are fully open then as contacts 1 to 7 apply ground power to the reset winding of K1. Resetting of K1 returns the circuit to its original condition; namely, drive power is removed from both the solid state transient switch and its associated logic circuit. The purpose of C1 will now be explained. When K2 is forced to reset by ground command, conditions would be proper for it to set itself, therefore attempting to nullify the ground command. At the start of the reset cycle C1 is discharged and must charge through R6. By making R6C1 time constant long enough, the base of Q5 will be held close to the negative bus in voltage, thereby preventing current flow through Q5 and hence through Q4. K2 is then prevented from attempting to set for a long enough time to allow K1 to reset which then removes power from K2 set circuit. At this point, the hybrid solid state switch is completely turned off.

From the foregoing circuit description, the operation of the hybrid solid state switch is seen to behave the same as a completely mechanical counterpart. Recent tests using the fabricated breadboard reveal the following data, which, at this time, is not yet fully analyzed for worst-case conditions. The data and conditions under which this information was obtained are as follows:

Initial preparation

Ground supply, 25 v DC
Spacecraft battery, 25 v DC
Load, 20 amp plus a 4000-μf capacitor
Cycle of operation = 2 to 5 sec per complete cycle

Results

Temperature, °C	Number of cycles of operation
−10	1000
+25, room temp	700
+75	1000

At this time, the design is being analyzed for worst-case conditions. Also component size reduction is being performed. An attempt will be made in the near future to package this design and to perform a *complete* environmental test in order to qualify it for flight availability.

V. Guidance and Control Analysis and Integration

A. Development of Advanced Spacecraft Operational Support Equipment

O. E. Linderman and R. V. Morris

The objectives of this operational support equipment (OSE) development task are to devise more comprehensive OSE testing techniques and other spacecraft tests which will result in a more reliably operating spacecraft. Subassembly or subsystem tests which will predict incipient failure of a spacecraft piece part or component are of primary interest.

The following testing techniques are currently being investigated: (1) margin test techniques for OSE and spacecraft equipment, (2) RF noise spectrums, and (3) infrared measurements of the various circuits.

A digital test circuit (*SPS* 37-26, Vol. IV, pp. 43–46) which cycles through a short operational routine is being used for evaluating the various testing techniques. Measurements are also being made on a spacecraft subsystem.

1. Voltage Margin Measurements

The voltage margin of a circuit can be defined as the amount the circuit supply voltage can vary from the nominal value before the circuit fails to operate properly. This voltage margin point is an important factor in circuit design and is generally specified in the requirements. The rate of change or the drift of this voltage margin point provides predictive information concerning future circuit operation. The object of this study is to find the correlation between the drift and other techniques for predicting circuit failure. Voltage margin measurements were made of the digital test circuit to study this test technique.

Margin measurements have been made on 29 of the 31 cards used in the digital test system. More precise and repeatable results were obtained by removing the circuit card from the system and making the measurements in a special test setup. A card mounted in this fixture can be seen inside the RF shielded box in Fig. 1. The +12 volt and −12 volt power supplies are varied to approximately 30% above and below the nominal 12 volt settings. The regions in which the circuit fails to operate are determined and plotted in a "shmoo" diagram, a two dimensional

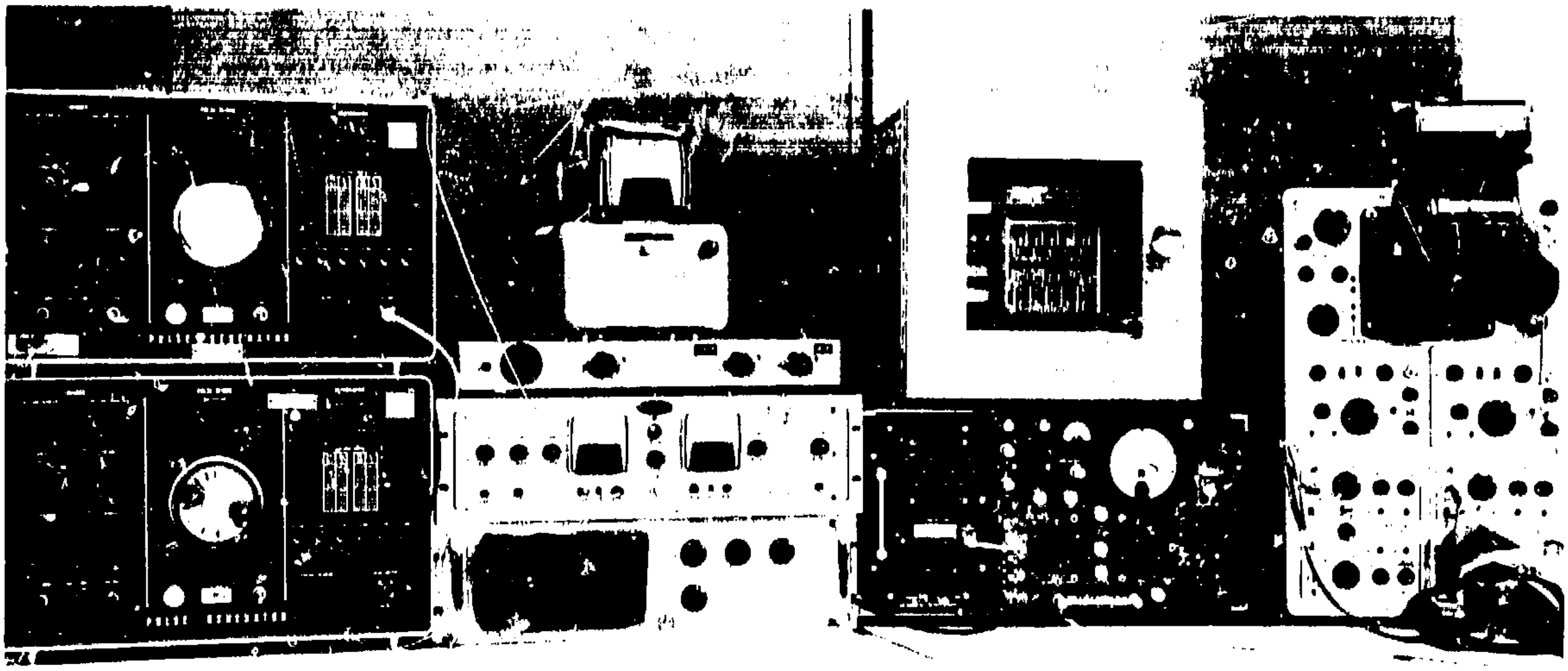


Fig. 1. Circuit card RF noise measurements

plot with +12 v supply voltage on one axis and −12 v supply voltage on the other.

Voltage margin measurements were also made on the complete digital test circuit. Fig. 2 is the system "shmoo" diagram showing various partial failures in circuit operation.

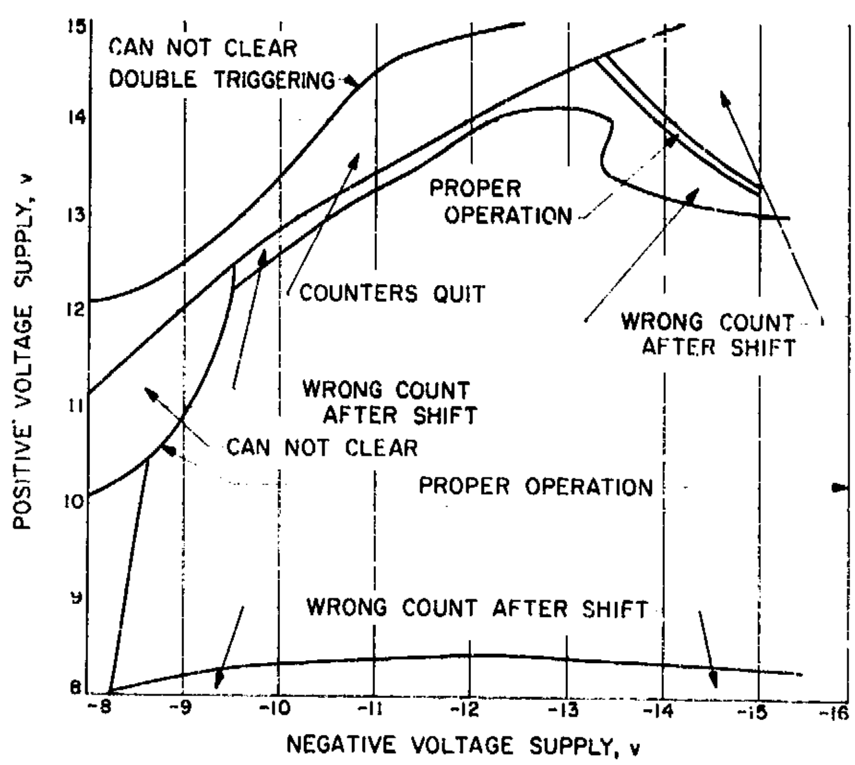


Fig. 2. Digital test system voltage margins

One of the limitations in using margin measurements for predicting future operation is masking. Malfunction of the weakest link in the circuit is the first to show up as the supply voltage is varied from nominal. At this point, the circuit may operate erratically or stop completely. Generally this malfunction can be traced to one voltage condition or a change in a piece part. If the circuit stops completely and only one margin is obtained, subsequent measurements show the drift only in this particular circuit margin or the variation in the one piece part. It may well be that another component is drifting much faster and will be the one to cause the circuit failure, but at the moment its margin characteristic is just outside the one presently measured.

Several margins for each circuit were obtained by using as many parameters or indications as possible for each circuit. The trigger and reset pulse amplitudes and pulse widths were varied when margin information could be obtained on a circuit. In cases where there are multiple circuits on a card, several margin characteristics were obtained for each circuit. These characteristics also were determined for different modes of circuit failure and have been plotted for ready reference. Voltage margin "shmoo" diagrams for a counter card are shown in Figs. 3 and 4. Fig. 3 shows the operation of the 5 counter circuits with a reset pulse that is 1.5 μsec long and has an amplitude of 5.3 v. Fig. 4 shows the operation of the counter circuit with a 1-μsec 2.65-v trigger pulse.

The voltage margin information obtained to date is voluminous, and has taken considerable time to collect.

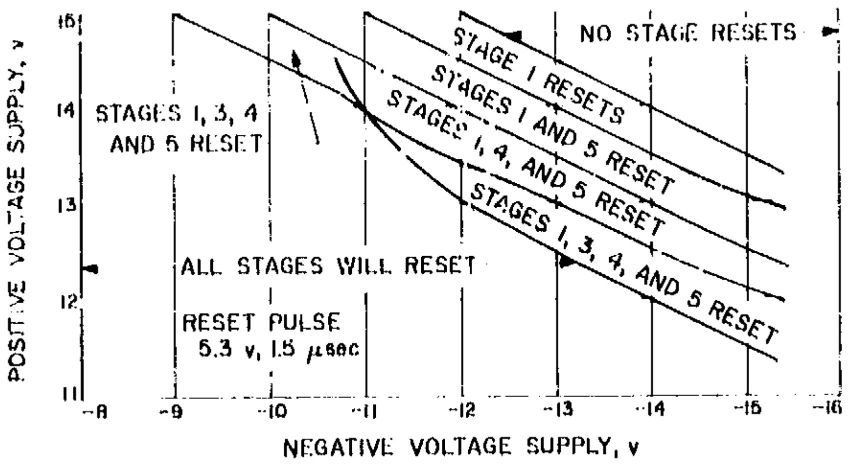


Fig. 3. Reset pulse voltage margins of counter card

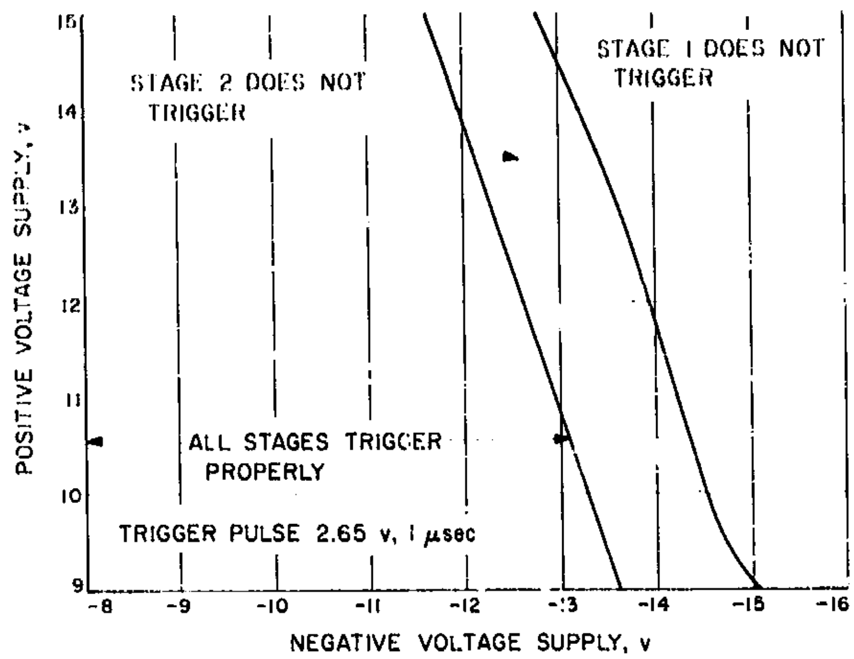


Fig. 4. Trigger pulse voltage margins of counter card

At the beginning of the project, some thought was given to more automatic methods of collecting the margin information. For example, this information could have been collected fairly easily during life testing of a developmental spacecraft CC&S subsystem controlled by a commercial Control Data 160A Computer. The tests were not run, however, after it was decided that the value of life-testing the particular type of CC&S construction did not justify the expense of obtaining the data.

Similar margin tests have been made by the Space Sciences Division on the *Mariner C* DAS spacecraft equipment using a commercial PDP-4 computer. Subsystem margin measurements were taken automatically for several months before launch on flight equipment and spares, and have continued on spare subsystems since. The results of this work are of great interest to the task described in this report.

While manual methods of obtaining the margin information have been time-consuming, it is felt the flexibility of this approach permitted various operational indications to be used which would be difficult to determine by automatic detectors. These various partial failure indications have provided additional margins of circuit operations. Having various margins for a circuit reduces margin masking and is expected to give additional information in case of circuit failure. The data that has been collected will now be used as a reference in future life test measurements of the circuit cards and the digital system.

2. Circuit Noise Measurements

The circuit noise of interest is composed of the random electrical fluctuations generated internally in electrical components. This electrical noise may be divided into five basic categories: (1) thermal noise, (2) shot noise, (3) excess noise, (4) avalanche noise, and (5) multistate noise. It has been recognized for some time that all electrical components generate electrical noise in one form or another. While the nature and amount of the noise varies with the type of component, all components contribute to the total noise of a circuit or system.

The noise measurement for a single component has been used as a factor to reject the component when the reading is abnormal. It has been determined that abnormal noise in a component is caused by some abnormality within the device and the offending component should be classified as a reliability risk. Commercial equipment is available (Ref. 1) for measuring the noise of many different type components, and these measurements are being made in present JPL component reliability studies.

The noise of interest in this study is the total circuit or system noise in the frequency range from 2 to 30 Mc. RF noise spectrums have been taken of individual circuit cards and of the complete digital test system using a recently purchased NF-105/M-126 receiver. The test setup for obtaining circuit card RF measurements is shown in Fig. 1. The receiver has excellent sensitivity and stability, and was designed as a standard for this type monitoring function. The receiver measures the RF energy or noise in a 10 kc bandwidth. The RF spectrum is obtained by tuning the receiver over the desired frequency band. The receiver indicates the RF energy in db on a meter, as well as giving an audio output in a headset.

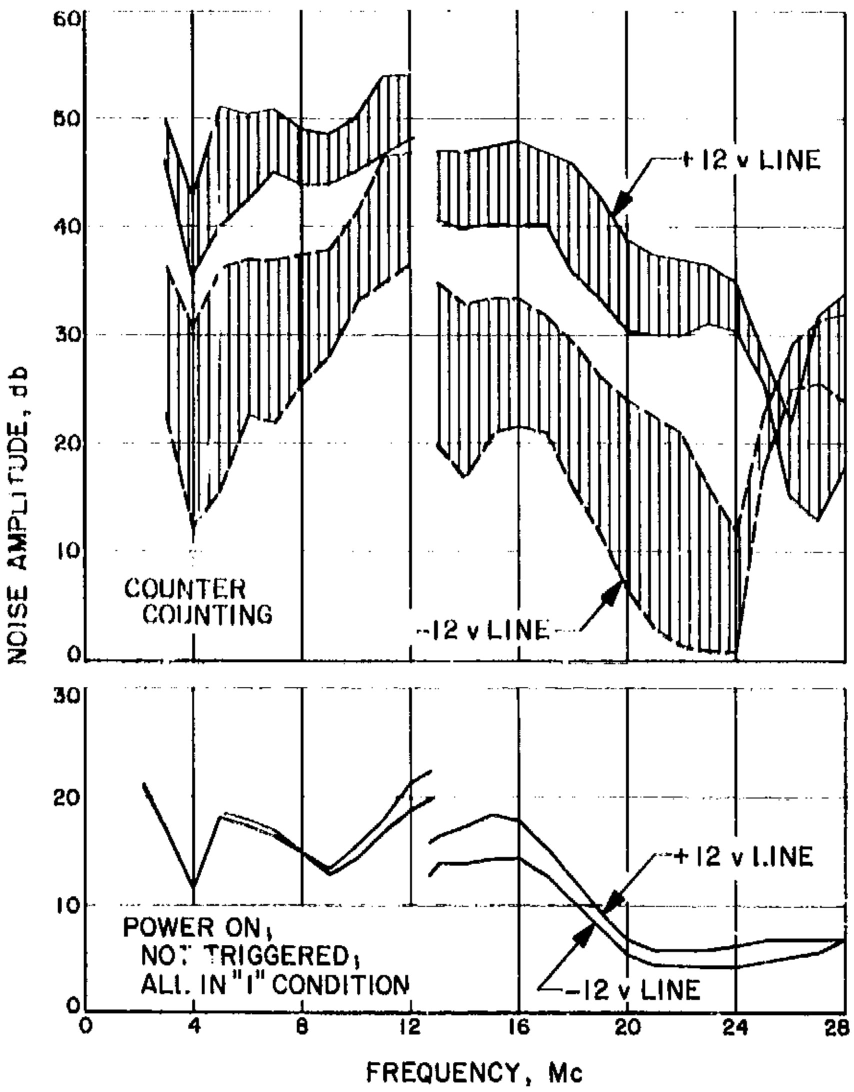


Fig. 5. Counter card 5 RF spectrums

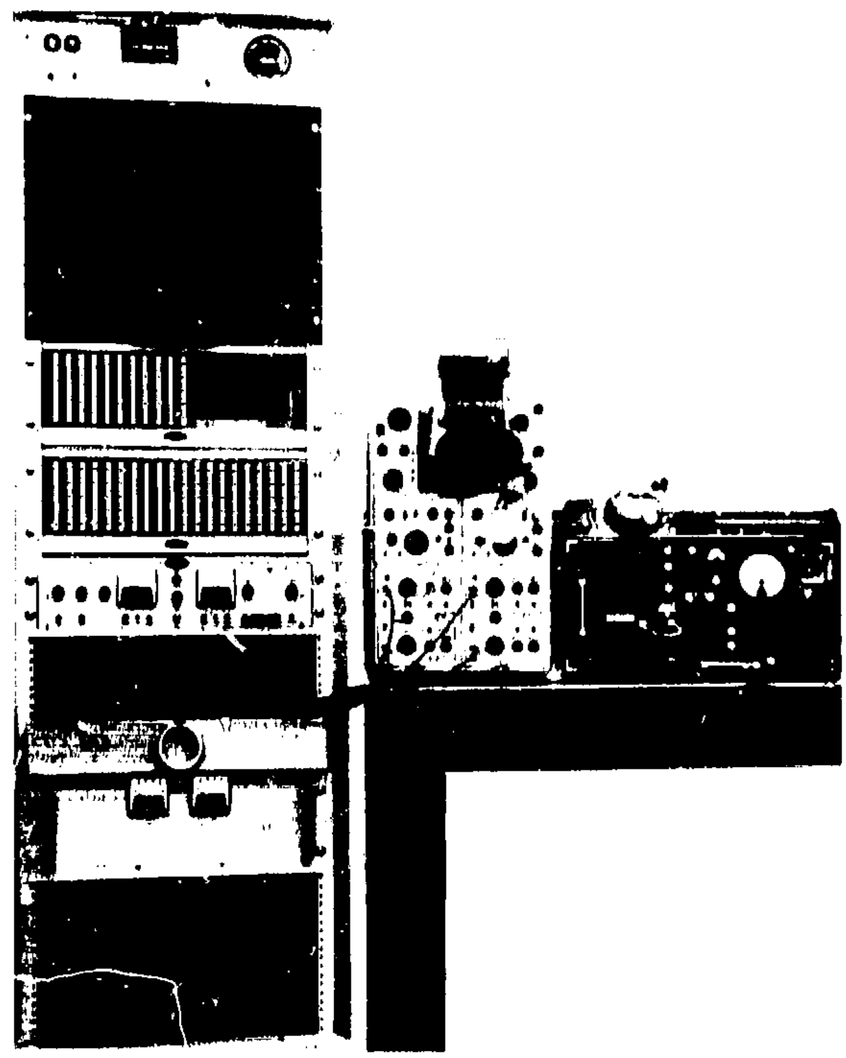


Fig. 6. Digital test circuit RF noise measurements

Preliminary RF measurements showed the need to shield the circuit card under test. The double-layer copper-screened box shown in Fig. 1 was constructed to hold the card and test fixture while obtaining the noise measurements. The measurements are made on the +12 v and −12 v power lines at the card, which are isolated from the power supplies by a filter which is additional to the feed-through filter into the shielded box. The noise voltage is taken from across a 50-Ω resistor in a ground return line of the filter, similar to the method used in RF tests of flight control systems (Ref. 2). The measurements were taken (1) with power on but the circuit not operating, and (2) with power on and the circuit in operation. Typical noise measurements of a 5 stage counter card are shown in Fig. 5. The digital operation gives the minimum-maximum readings shown by envelope lines in the upper part of the figure. The discontinuity at 12.7 Mc is due to changing receiver tuning bands. No attempt was made to adjust to a common point as the readings are relative.

The digital test system was mounted in a partially enclosed rack (Fig. 6) for the RF measurements. Some of the system RF Spectra are shown in Fig. 7. Only the +12 v line noise (when the circuit is in operation) is shown in order to reduce confusion in the regions where the +12 v and −12 v noise graphs overlap. The discontinuities at 12.7 Mc and 5.2 Mc again are due to receiver band changes. The system noise with power on but the clock stopped was quite high, so these values are calibrated relative to a 32-db reading on the −12 v line at 28 Mc when the test circuit is in operation.

RF noise measurements were also taken on a *Ranger* Central Computer and Sequencer (CC&S) operating with its subsystem OSE (see Fig. 8). The preliminary readings were slightly erratic. The receiver was later found to be intermittent on some bands, and has now been returned for servicing and repair. The amount of interference on the monitored lines indicates that the OSE, CC&S, and cables may have to be placed in a large RF-shielded room.

As yet there has not been an indication that the noise of a defective component will be easily recognized by the present measurements, although others have had some success in using RF noise to locate defective components

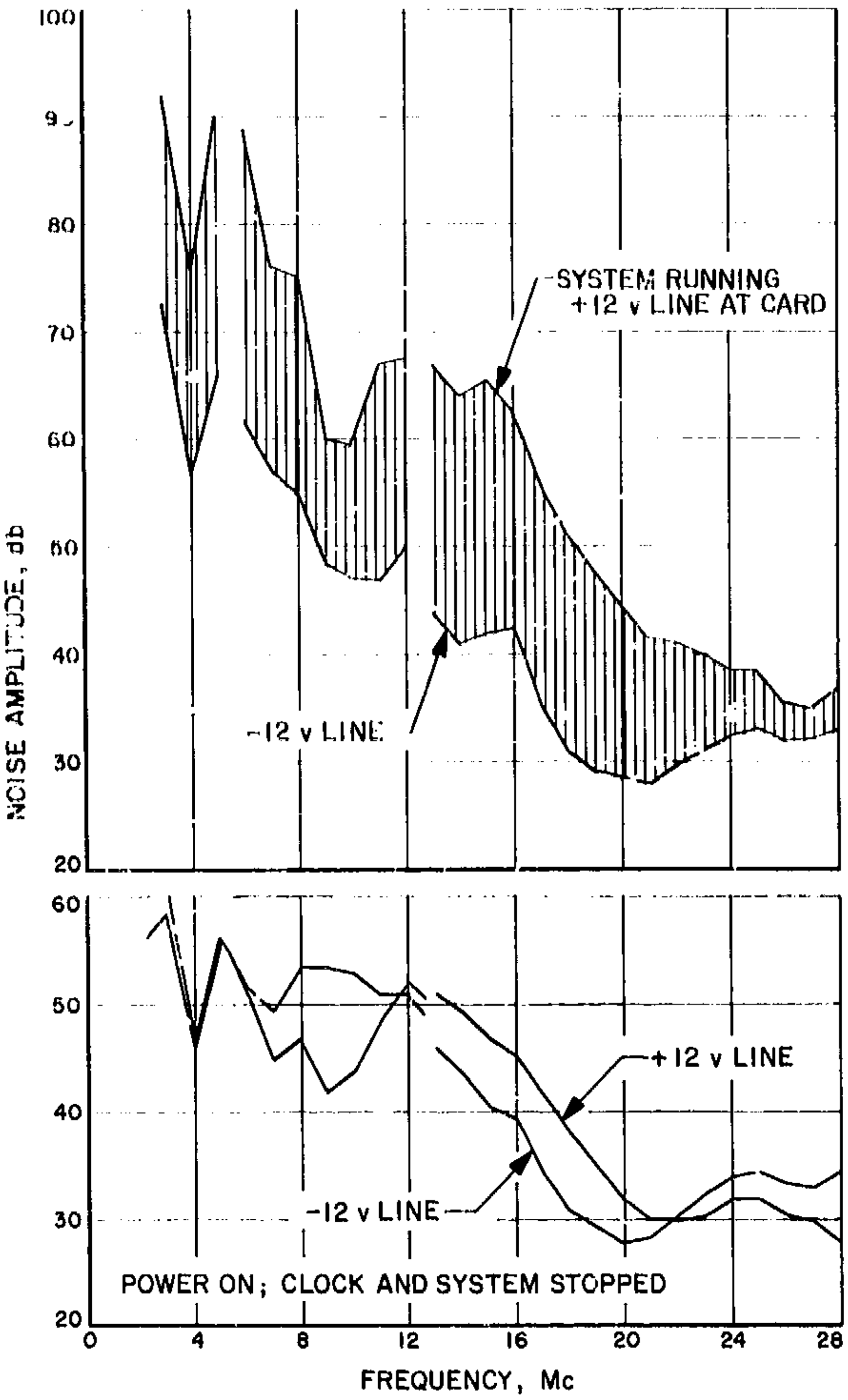


Fig. 7. Digital test system RF spectrums

in control systems. However, these were circuits with many stages of gain and an output response that was continuous and fairly linear. In contrast, the individual circuit cards discussed here do not have many components per card, and the circuit operation is almost completely digital. The noise for the power-on condition, but where there is no digital triggering, measures just slightly above the receiver threshold in most cases. When the circuit card is triggered in digital operation, the signal interference causes large amplitude changes or beating. While the various counters, registers, and gates have noise characteristics or spectrums which easily distinguish each circuit, more measurements will be made to determine repeatability in measurements. The worth of the RF measurements in this digital application will be determined in future life tests.

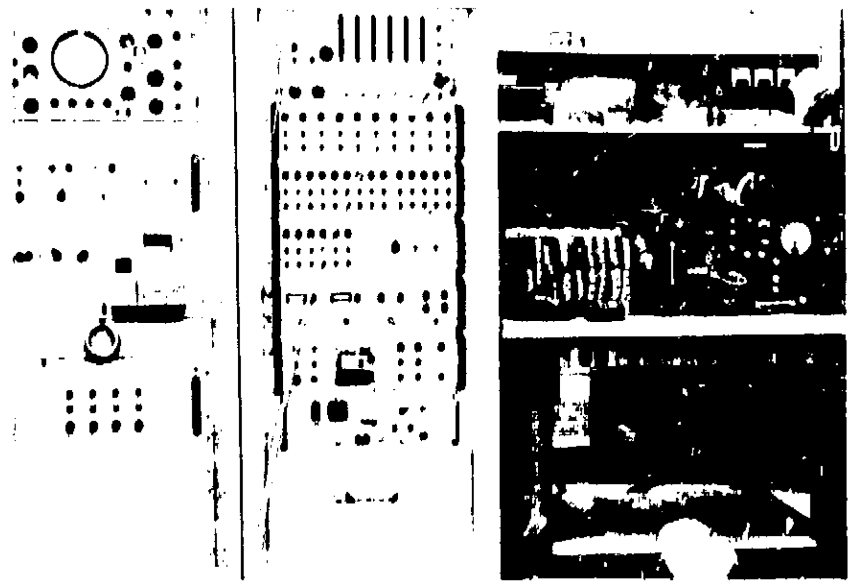


Fig. 8. *Ranger* CC&S RF noise measurements

3. Infrared Measurements

Very few infrared measurements have been taken of circuit cards since the last report. A new infrared camera was purchased last Fall, so measurements had been postponed so they might all be taken with the same camera. This camera is a Barnes Engineering Model T-4D and has a 4-to-1 improvement in spatial resolution over the Model 1-8A used previously. Unfortunately, the improved resolution was obtained at the expense of sensitivity. Part of the increase in resolution was obtained by increasing the bandwidth of the radiometer detector electronics. The resulting noise gives a very noisy background in the present photos for temperatures slightly above the ambient room levels. The photos therefore somewhat limit the infrared data that we desire to show. Possible bandwidth modification, substitution of detectors, and decreasing the scanning time are being studied to obtain the best solution to present limitations.

It has proved quite difficult to accurately measure the point-to-point temperature on a circuit card even with the best photographs. Such temperatures are determined by interpreting the gray scale of the photo with that of six calibrated temperature standards placed so they are also photographed. Individual component temperature measurements have been obtained by stopping the camera from scanning and focusing on just one component. The voltage output (or temperature) is recorded as well as that from the calibrated standards which bracket the reading. This means of taking measurements is slow and cumbersome.

One company has automated (Ref. 3) the taking of individual infrared detector voltage measurements with the use of an automatically controlled mounting surface which can locate the circuit card under test in two dimensions to 0.001 in. The circuit card is positioned, the reading taken, and the card moved to the next component position automatically. These voltage readings are compared against stored information, and any changes in values are read out. This equipment was placed in operation late last Fall; calibration runs and the elimination of air currents are continuing in the evaluation of the infrared measurements.

Such an infrared test station would greatly speed the recording of individual component temperature measurements that we are making on circuit cards. The development of this test station is of great interest to this task for possible future application.

An infrared radiometric microscope is also being considered to obtain the individual component temperature measurements. The microscope can be focused quite rapidly and accurately on an individual component; however, the temperature data would be manually recorded. Thus it is a time-consuming process to determine the general overall circuit card heat distribution with the scanning camera, use the radiometric microscope to obtain more accurate temperature data of selected components, and manually recording of data. The automation of any steps in the process would be of great aid in obtaining the data, storing it, and using it in the future.

The present effort is concentrated on a complete evaluation of the scanning camera. The circuit card photo remains the chief means to quickly evaluate the overall circuit for thermal operation. Means to obtain more accurate data by automatic means will continue to be considered.

References

1. Curtis, J. G., *Current Noise Tests Indicate Resistor Quality*, Quan-Tech Laboratories, Inc., Whippany, N. J., January 1962.

2. Maki, C. E., et al, *Detection of Electrical Faults by RF Techniques*, Honeywell Aeronautical Division, Minneapolis, Minn., October 1963.

3. Judd, L. R., and Magee, T. J., *Tolerance Studies for Infrared Production Testing of Electronics*, The Boeing Company, Huntsville, Ala., February 24, 1956.

VI. Guidance and Control Research

A. Magnetics Research

C. H. Wilts[1]

1. Ferromagnetic Hysteresis Instrumentation

The magnetic characteristics of thin films can be measured with a number of instruments. One of the simplest of these is the hysteresis loop tracer (Refs. 1, 2) which permits direct measurement of saturation magnetization, remanent magnetization, coercive force and hysteresis loss due to a unidirectional oscillating magnetic field. In addition, the anisotropy of a film can frequently, though not always, be measured with this instrument. Most instruments of this type measure the component of flux in a given direction by means of the electrical signal induced in a loop of wire, although some instruments have been described in which an optical measurement is made using the Kerr-magneto-optic effect (Ref. 3) or the Faraday effect (Ref. 4).

The results of measurements with this type of instrument depend to some extent on the frequency of the oscillating field. Ideally, one would like to measure the magnetic characteristics mentioned above at a frequency so low that they are independent of frequency. Because of instrument limitations and, in some cases, because of film characteristics, this means that measurement should be made below 60 cps, and for some films perhaps below 20 cps. The present instrument retains the sensitivity of earlier instruments operating at 1000 or 60 cps, but permits measurements anywhere in the frequency range from 0.1 cps to 100 cps.

2. Instrument Construction

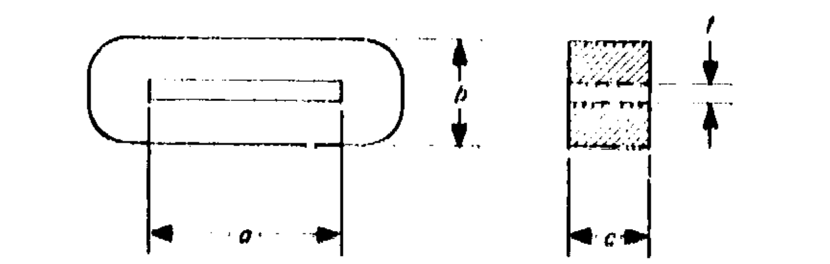
a. Integrator. The critical components of the instrument are the active integrator and the pickup loop. There are many operational amplifiers available that can be used for the integrator. Considering both gain and noise the Kintel 111BFO or 112A appears to be superior to other available amplifiers by a factor of 2 or more. It is not practicable to design one pickup loop and one integrating circuit to cover the three-decade range of frequency from 0.1 to 100 cps. However, operation over a range wider than one decade is easily accomplished. This article will describe two such circuits, one for operation below 10 cps and the other for use between 10 and 100 cps.

b. Low-frequency integration. Details of the pickup coil are shown in Table 1. The coil contains 15,000 turns of No. 52 enameled copper wire. This wire has a diameter of about 0.0008 in. and a resistance of about 1.5 Ω/in. The

[1] At the California Institute of Technology, performing work supported by the Jet Propulsion Laboratory.

Table 1. Pickup coil details

Parameter	Low frequency	High frequency
a, in.	1.10	1.10
b, in.	0.25	0.25
c, in.	0.25	0.15
t, in.	0.06	0.05
Number of turns	15,000	7,000
Wire size	52	48
Resistance, Ω	60,000	11,000
Inductance, h	3.1	0.80
Q (1000 cps)	0.32	0.46



wire is easily broken and great care must be used in winding the coil. When wound to the dimensions shown in Table 1, the coil will have a resistance of about 60,000 Ω, an inductance of about 3 H, and a resonant frequency (from interturn capacitance) of about 10^4 cps. When used as an integrator, the resistance of the coil serves as the input resistor and the feedback capacitor can be chosen to give the desired sensitivity. Choice of capacitor affects the signal-to-noise ratio as well as the sensitivity. In principle, reducing capacitance suppresses high-frequency noise relative to the signal (leaving low-frequency noise to signal unchanged), and so should improve the signal-to-noise ratio. However, for capacitance below about 0.01 μf there is no significant improvement since the frequency band affected contributes negligible noise.

The circuit for the integrator is completely conventional, but for completeness is shown in Fig. 1(a). When

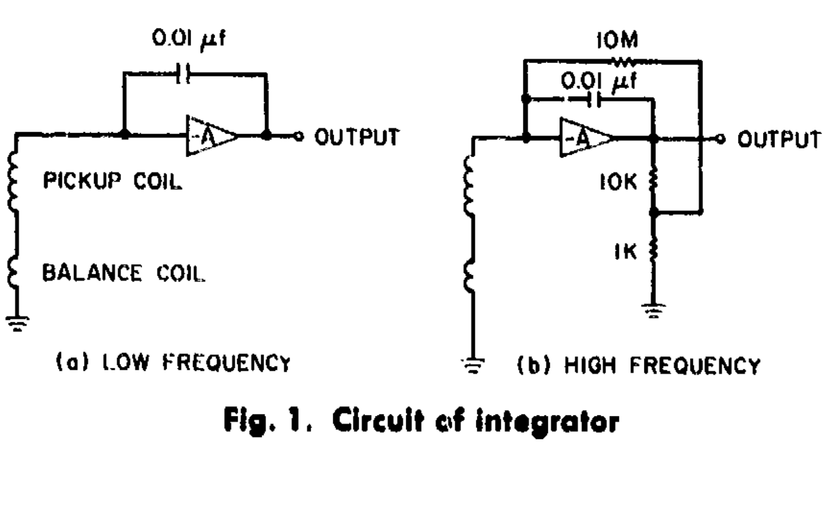


Fig. 1. Circuit of integrator

drift has become excessive, the integrator may be reset to zero by providing a shorting switch across the integrating capacitor, or the same thing may be accomplished by merely rotating the gain knob on the amplifier from the operational position to the gain of 20 position. The balance control on the amplifier is much too coarse, but its full range is needed since vacuum-tube characteristics and bias cells in the amplifier change with time. The simplest solution is to replace the balance potentiometer with a small 10 K, 10-turn potentiometer (type B Ford Multipot is suitable) which requires minor relocation of components on the back side of the front panel.

A final point requires consideration for use at low frequencies. Bakelite or even Lucite provides inadequate insulation for the input circuit when the 60,000-Ω input circuit is used. Consequently, the plastic structure holding the pickup coil is mounted on thin grounded washers, and signal leads to the integrator are all in grounded shields.

c. High-frequency integration. For higher frequency use, a pickup coil with fewer turns may be used. To compensate for the reduction in turns, the integrator gain may be increased by reducing the input resistance or the feedback capacitor. The signal-to-noise ratio may be maintained at a favorable value by reducing the low-frequency noise and drift by use of a suitable large resistor across the integrating capacitor. Details are given in Table 1 for the pickup coil used between 10 and 100 cps. The inductance limits the maximum allowable frequency to about 100 cps. To reduce the low-frequency drift and noise, the capacitor should be shunted by a resistor to give a time constant of approximately 1 sec. With capacitance of 0.01 μf, the required resistor is inconveniently large so the equivalent network shown in Fig. 1(b) may be used. This resistance network determines the lower frequency limit of the instrument. The phase shift of the integrator differs from 90 deg by about 1 deg at 10 cps, and the deviation increases by a factor of 2 for every decade decrease in frequency below this value. No capacitor shorting switch is necessary when used with this network, although the balance control should be used to keep the dc level of the output reasonably close to zero.

d. Auxiliary components. Balance coils are needed to cancel the voltage induced in the pickup coil when no magnetic sample is in place. The geometry and wire size are relatively unimportant. For the instrument described, a fixed coil of about 1000 turns of No. 34 wire on a 2-in.

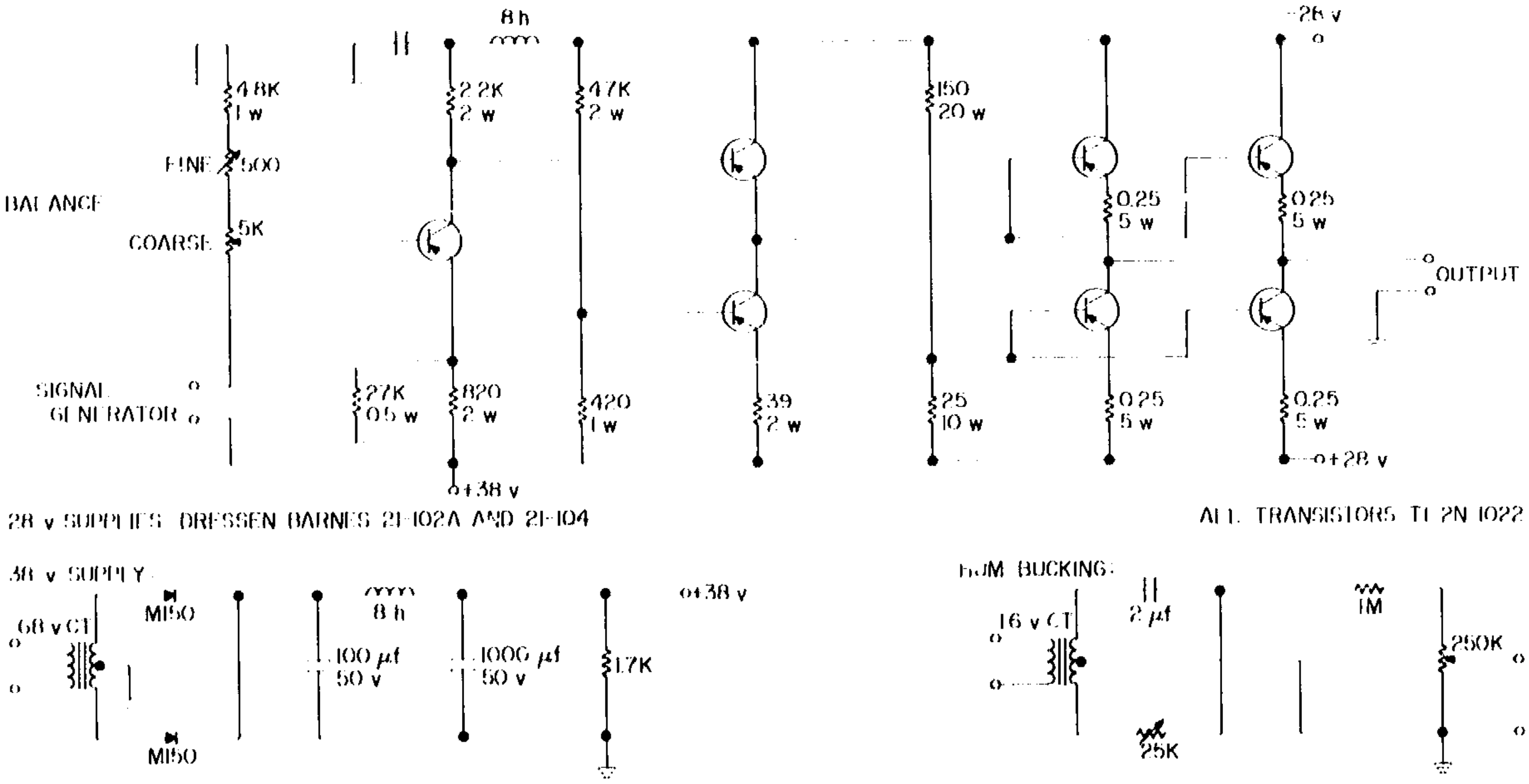


Fig. 2. Transistor driver for Helmholtz coils

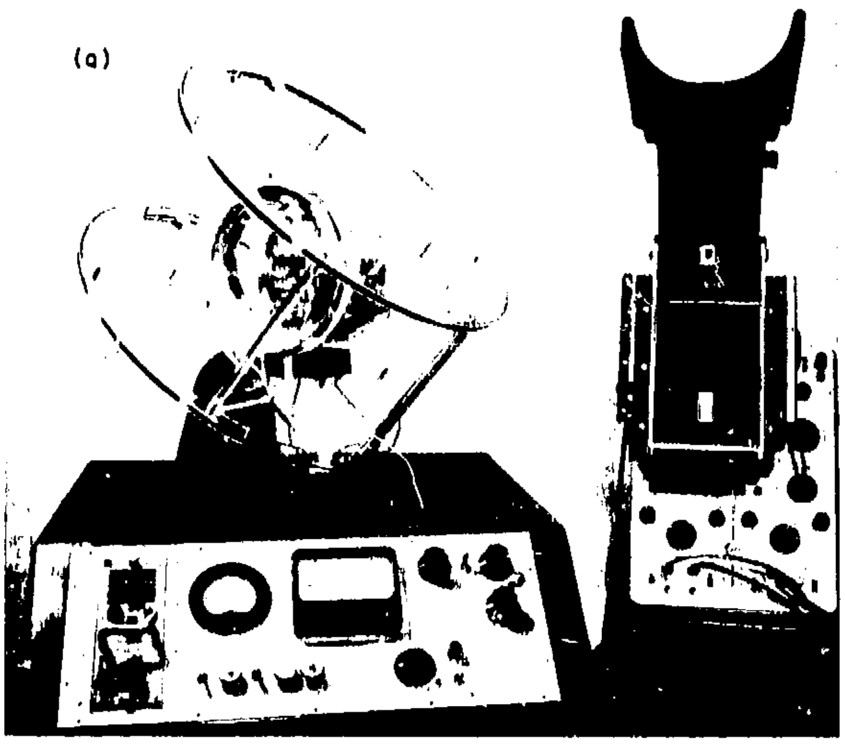
diameter spool is used. An additional movable coil of half the diameter and with about 25 turns is used for final adjustment of the balance which drifts slightly over long periods of time because of such things as thermal expansion of components.

The magnetic field for the instrument is provided by Helmholtz coils with mean diameter 7 in. and mean spacing 3.5 in. Each coil has 550 turns of No. 20 wire providing a field of 100 Oe with 1.8 amp at 40 v for series connection, or 3.6 amp at 20 v for parallel connection. The power amplifier to drive these coils must operate from DC to 100 cps. A simple class AB direct-coupled transistor driver is shown in Fig. 2. A more sophisticated design, or higher power components, would permit fields in excess of 100 Oe. The power amplifier is, in turn, driven by a signal generator at the desired frequency. Particularly at low frequencies, it is desirable to vary the field at a uniform rate, so an approximately triangular wave form is desirable. The Hewlett-Packard function generator Model 202A is well suited to this application. The magnetic field is metered by a resistor in series with the drive coils. For the parallel connection and the particular geometry used, a resistor of 0.270 Ω gives a signal of 10 mv per Oe. Details of the head and the arrangement of the drive coils can be seen in Fig. 3(b).

The most critical problem in the design of the drive coils is to ensure that the magnetic field at the sample is precisely in phase with the current in the coils. To avoid compensating devices or circuits, it is desirable to avoid all metal parts in the vicinity of the drive coils. Eddy currents in such parts set up perturbing fields at the same drive frequency but with unpredictable phase. In the present instrument, plastic parts and Nylon screws are used throughout the critical region.

The Earth's field is cancelled by a large pair of Helmholtz coils 20 in. in diameter. Each coil has 200 turns of No. 28 wire, requiring a direct current of about 0.05 amp at 6.5 v. If there is magnetic coupling between these coils and the drive coils, then induced currents in these coils can produce an extraneous field at the drive frequency. To eliminate this, the Earth's field coils are mounted with axis perpendicular to the axis of the drive coils as shown in Fig. 3(a).

All AC operated equipment has some 60-cps pickup. In this apparatus most of the unwanted 60-cps signal arises from the power supply transformers of the signal generator and integrating amplifier. This was minimized by moving the transformer of the generator to a remote location and by providing at the oscilloscope input a small 60-cps signal adjustable in phase and magnitude.



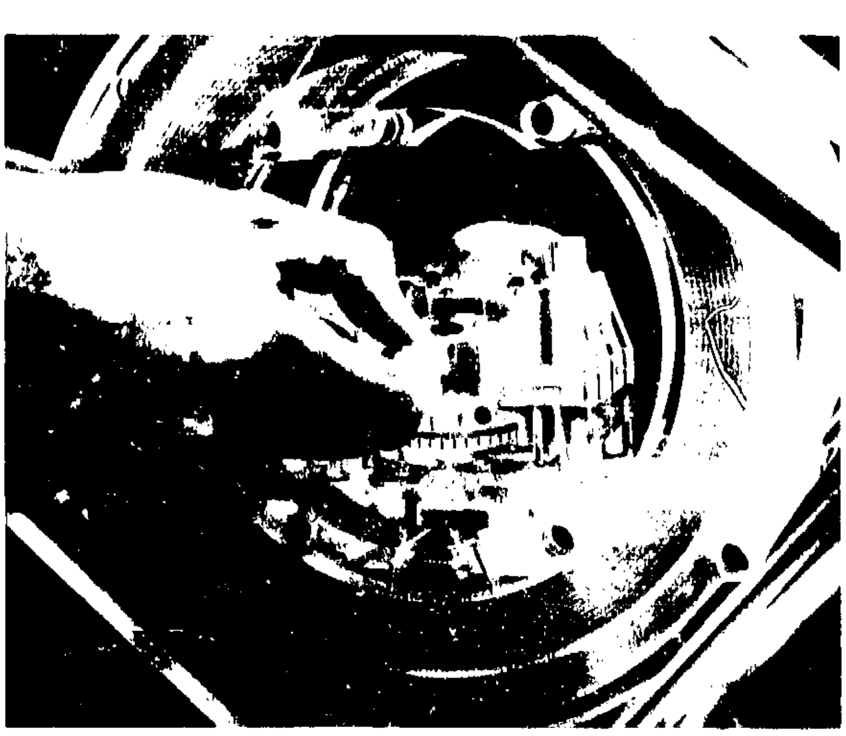


Fig. 3. Hysteresis loop tracer instrument

When properly adjusted, the residual signal of about 200 μv consists of harmonics of the power frequency.

Observation of hysteresis loops is made using a Tektronik 503 oscilloscope. Since the internal calibration of the oscilloscope does not have the accuracy desired of this instrument, an auxiliary calibrator (better than 0.5%) is provided which can be applied to either horizontal or vertical deflection inputs. With such a calibrator, measurements of saturation magnetization, coercive force, and anisotropy field H_k may be made routinely with 1 to 2% accuracy.

The oscilloscope, calibrator and integrating amplifier should be operated from a regulated AC line. However, the drive amplifier should not be operated from this line due to interaction at low frequencies with the integrator through the power supplies. It may, however, be connected to the AC line on the other side of the regulator without appreciable interaction.

3. Instrument Performance

The hysteresis loop tracer described here is ordinarily used to measure magnetic properties of thin films. The output signal is directly proportional to thickness and width of the sample, and more or less independent of length if greater than 1 cm. For a sample 0.75 cm long the response is reduced 10%, and for a sample 0.5 cm long it is reduced 30%. In what follows the response will be referred to a sample 1 cm square. With the low-frequency (DC) integrator, basic sensitivity of the instrument is about 200 mv per Maxwell (gauss cm^2). Referred to a 1-cm wide sample and in more common units, this is 500,000 gauss angstroms per millivolt. Since the noise level of the instrument is approximately 0.3 mv, the signal is equal to the noise for a sample of about 150,000 gauss angstroms, or a sample of permalloy about 15 Å thick. In other words, if the coercive force is under 100 Oe, ferromagnetism can be detected in a sample about 10 Å thick; but for reasonably accurate measurement of magnetic properties, the thickness of the sample should be not less than 50 Å.

For the high-frequency integrator, the performance is very similar. The number of turns on the pick-up coil is reduced by a factor of 2 while the integrator gain is increased by a factor of 6. The over-all sensitivity is therefore increased by a factor of about 3. Since the resistive feedback reduces the low-frequency components of noise, the over-all signal-to-noise ratio is nearly unchanged.

Typical records of magnetic characteristics are shown in Figs. 4 to 6. In Fig. 4, the sample is 75 permalloy about 2000 Å thick. Slow flux reversal in the easy direction is by jerky wall motion with corresponding Barkhausen jumps in the hysteresis loop. These show up very well at frequencies below 2 cps but disappear above 10 cps because of the high-frequency phase shift and attenuation caused by the inductance of the pickup loop. The response of the high-frequency integrator is shown in Fig. 5 for the same sample. At 10 cps the low-frequency phase shift is observable in the displacement of the two traces in the horizontal part of the hysteresis loop. At

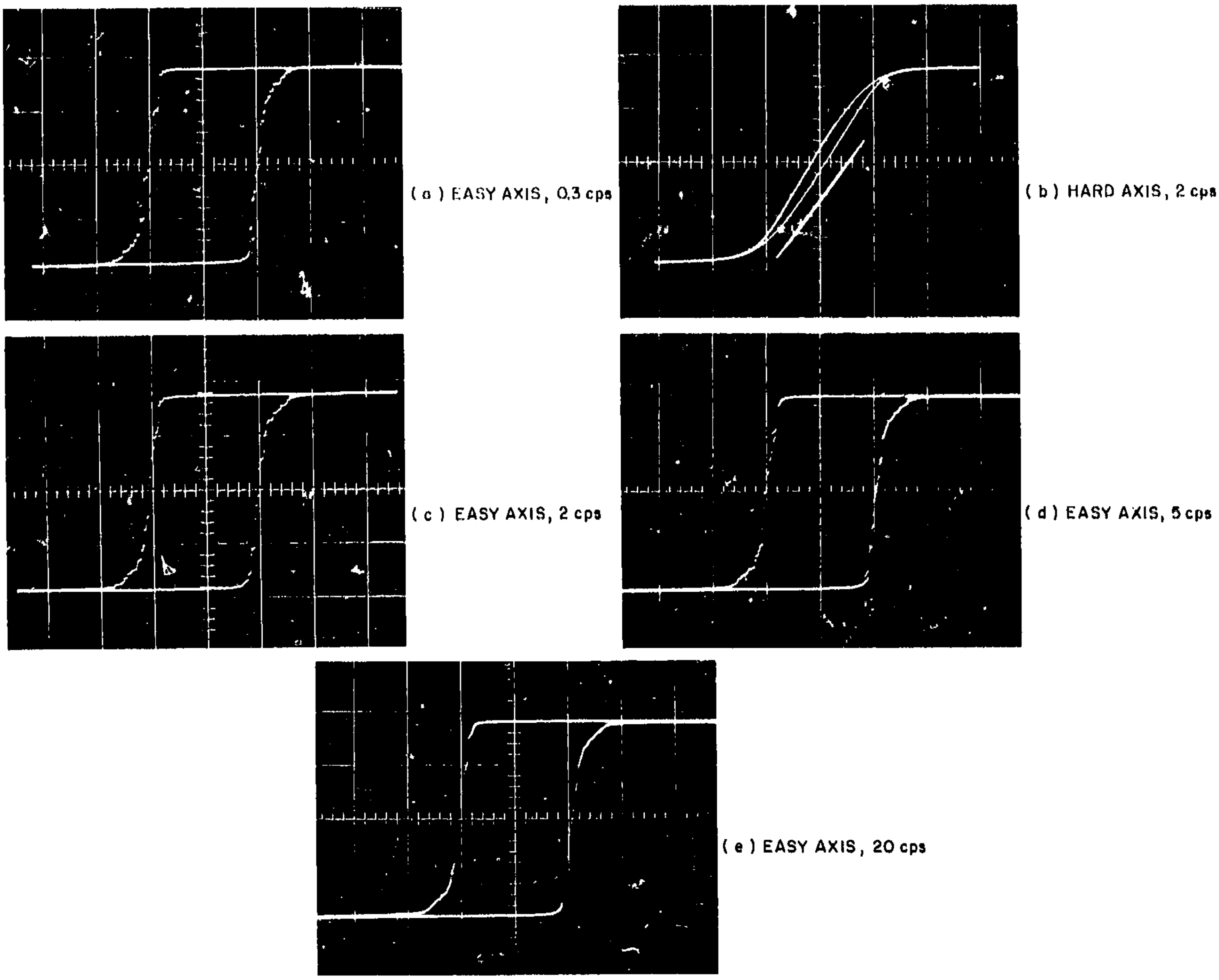


Fig. 4. Hysteresis loops of a ferromagnetic film with low-frequency integrator; horizontal scale 2 Oe/cm easy axis; 10 Oe/cm hard axis

20 cps, this phase shift is negligible but the high-frequency detail is beginning to disappear in the wall motion portion of the loop. At 50 and 100 cps, progressively increasing high-frequency phase shift alters the apparent coercive force and even more seriously causes an apparent change in coercive force with the amplitude of the driving field. As far as the hard axis loop is concerned, there is no significant difference in performance in the same frequency range. Consequently, only one example of hard axis hysteresis loop is shown in Fig. 4. Performance of the high-frequency integrator can be improved slightly by changing integrating resistor to 100,000 Ω and integrating capacitor to 0.001 μf (shunt resistor is in turn increased to 1000 MΩ), but the improvement is not considered worth the trouble of changing the feedback capacitor which is then required when changing pickup coils from low- to high-frequency integration.

Typical records with a very thin film are shown in Fig. 6. The sample is 80 permalloy about 50 Å thick. The coercive force can be measured with good accuracy (1 or 2%), but saturation magnetization and anisotropy field are uncertain to about 5 or 10%.

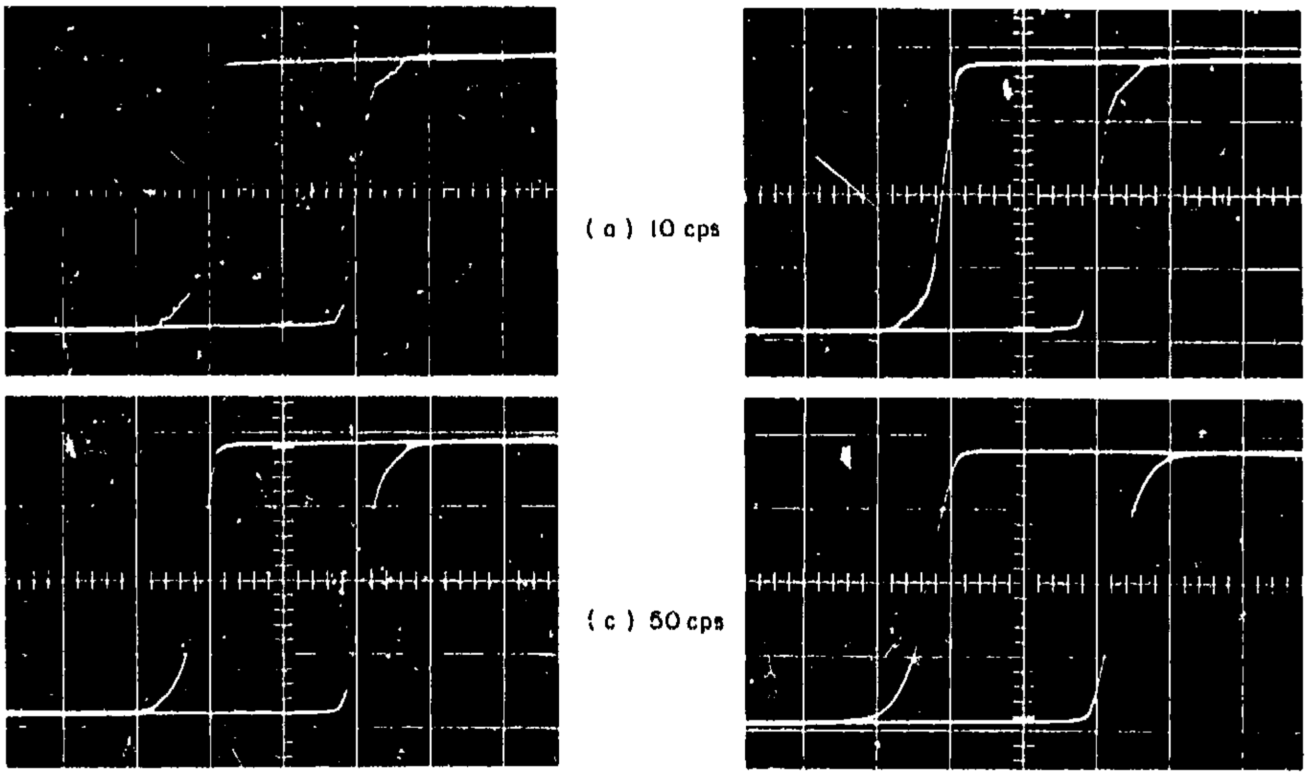


Fig. 5. Hysteresis loops with high-frequency integrator; horizontal scale 2 Oe/cm

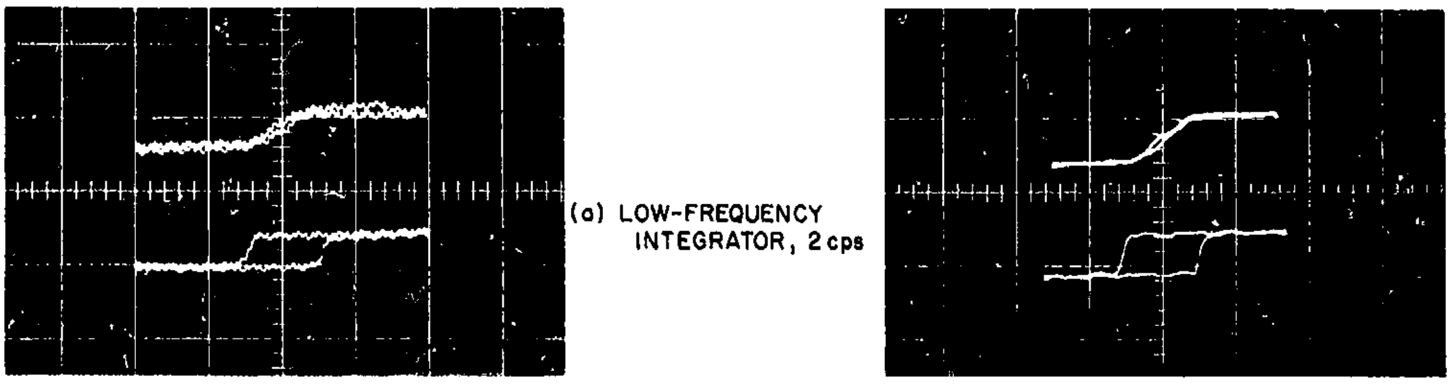


Fig. 6. Hysteresis loops of 50 Å permalloy film; horizontal scale 2 Oe/cm easy axis; 10 Oe/cm hard axis

For a detailed record of easy axis switching, the low-frequency mode of integration is preferable. However, there are two major disadvantages:

(1) Accurate visual observation is difficult even with a long persistence screen on the oscilloscope. This means a photographic record is almost a requirement for every measurement.

(2) Because the gain of the integrator has been raised to the maximum useful value, drift of the integrator is a serious problem for films less than 1000 Å thick. Continual adjustment of the balance control is required.

Because of these disadvantages, experience has shown the high-frequency mode to be preferable for routine measurements. For best accuracy consistent with ease of operation a frequency of about 20 cps seems most satisfactory. In actual practice, 17 cps is used to avoid a subharmonic of the line frequency, although this precaution is only important for very thin magnetic samples.

Analysis of performance, frequency limitations, and signal-to-noise ratio are facilitated by use of the equations given below.

a. Low frequency ($R_0 = \infty$)

$$E_0 = \frac{E_s}{\left[RCs\left(1+\frac{L_s}{R}\right)+\frac{1}{A_0}\right]} + \Delta E\left(\frac{RCs+1}{RCs}\right) + \Delta I\left(\frac{1}{Cs}\right)$$

b. High frequency

$$E_o = \frac{E_s}{(RCs)\left(1 + \frac{Ls}{R}\right)\left(1 + \frac{1}{R_o Cs}\right)} + \Delta E\left(\frac{RCs + 1}{RCs + \frac{R}{R_o}}\right) + \Delta I \frac{1}{Cs\left(1 + \frac{1}{R_o Cs}\right)}$$

The schematic circuit with all important elements is shown in Fig. 7. However, it should be noted that the elements R_i R_g and C_i have no appreciable effect in the frequency range of interest. Almost all limitations are imposed by L, R_o and ΔE.

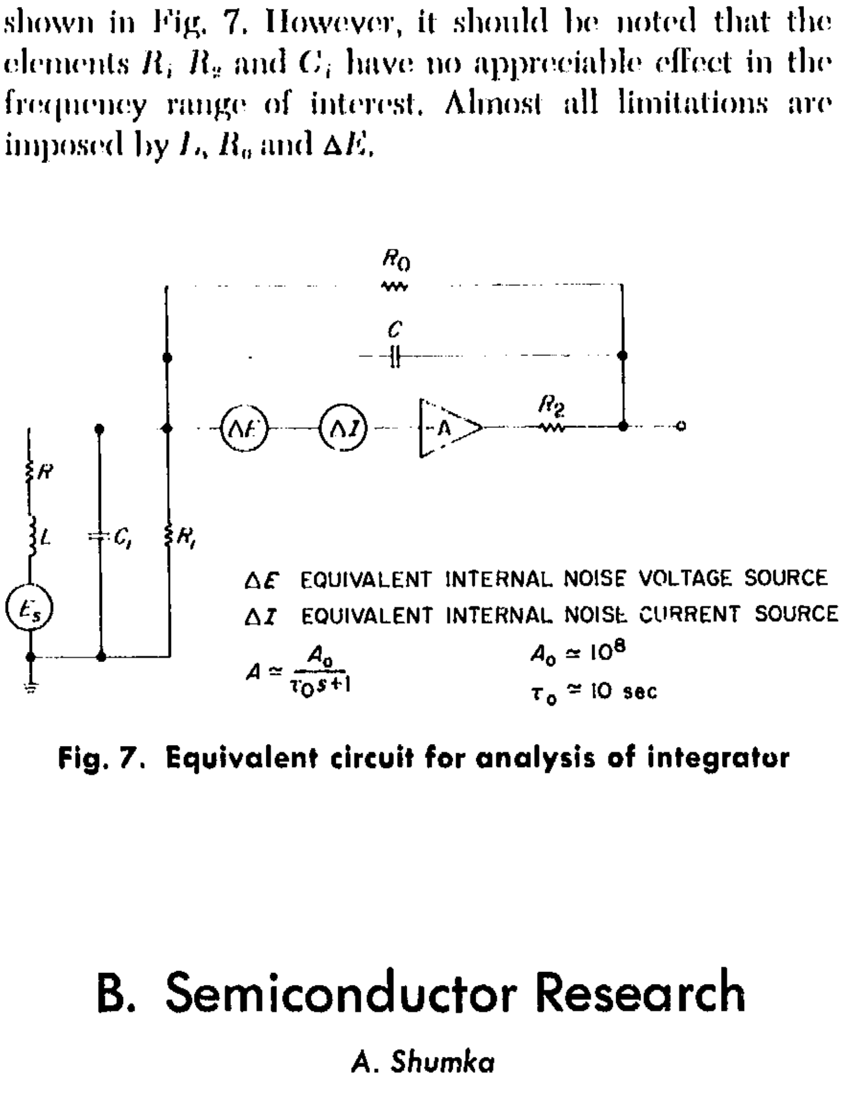


Fig. 7. Equivalent circuit for analysis of integrator

B. Semiconductor Research

A. Shumka

1. Space-Charge-Limited Current in Germanium: Fabrication of Solid-State Diodes

a. Introduction. Space-charge in homogeneous semiconductors and insulators may arise from charge accumulation, charge depletion and/or charge trapping. Each of these influences the current conduction in a specific manner. This unique relationship between the type of space-charge and the current conduction provides a very useful method for investigating space-charge effects. It has been successfully applied in determining the properties of traps in such materials as CdS (Ref. 5). Another important application of this technique is to investigate charge accumulation in homogeneous semiconductors and its effect on limiting the flow of current.

Space-charge in solids due to the accumulation of conduction carriers establishes an electric field which tends to limit the flow of current. Limitation in current flow resulting from this effect only is defined here as pure space-charge-limited (scl) current, and is analogous to that which exists in thermionic vacuum-tube diodes. Only a small number of experiments have been performed to investigate pure scl current in practical solids. The initial attempt to observe pure scl current in solids was made by Dacey (Ref. 6), whereby it was indicated in a semiquantitative manner that pure scl hole current can flow in germanium. Wright (Ref. 7) made similar measurements on CdS and observed at high current levels a square law dependence between current and voltage. This relationship is expected from the theory for field independent mobility. Recently the author (Ref. 8) carried out extensive measurements on pure scl hole current in high-purity near-intrinsic n-type germanium (ν-type). From these results a direct correspondence between the measurements and the theory for field-dependent mobility (Ref. 6) was obtained.

Experimental research work on space-charge effects in homogeneous semiconductors was started at JPL in September 1964. The objective of the project is to investigate pure scl current flow in a solid using near-intrinsic high-purity germanium as the conduction medium. This material is suitable for studying the transport mechanisms of electrons and holes for scl current flow. Previous work has experimentally shown (Ref. 8) that pure scl hole current can flow in germanium. A similar experiment is planned to study pure scl electron current in germanium. From the combined results a more detailed understanding of the transport properties of electrons and holes can be obtained. Furthermore, these results will contribute toward demonstrating the feasibility of such electronic devices as the solid-state diode and related devices which operate in the scl current mode.

b. Solid-state diodes. Near-intrinsic high-purity p-type (π-type) with an acceptor doping density of the order of 10^{13} cm^{-3} is used as the conduction medium. A useful experimental structure consists of a thin π-type germanium wafer bounded by two electrical contacts of which one is an injection contact and the other a collector contact. This is analogous to the geometry of a planar thermionic vacuum-tube diode with the cathode, anode and vacuum corresponding to the injection contact, collector contact, and germanium wafer, and is called a solid-state diode. Two highly doped n^+-type regions are alloyed onto the germanium wafer to form the necessary electrical contacts. The solid-state diode as shown in Fig. 8(a) resembles a transistor with a floating base configuration. Transistor terminology where applicable will be used, as is evident in Fig. 8(a).

Only a brief outline of the operation of the solid-state diode will be presented. A more comprehensive explanation is given in Ref. 8. When a small voltage (V) is applied, the emitter becomes forward-biased and the collector reverse-biased. The charge-depleted region of the collector, across which most of the applied voltage appears, extends from the collector junction into the base. Subsequent increases in the applied voltage will cause the collector depletion region to widen until eventually the base region is completely charge-depleted. The lowest applied voltage at which this occurs is called the punch-through voltage V_{pt} and, at this point, the solid-state diode loses its identity as a transistor with a floating base. For applied voltages slightly larger than V_{pt}, the current increases rapidly because the additional electric field can be sustained only by charge carriers (electrons) injected into the base from the emitter. As long as the electron density is small compared with that of the fixed ionized acceptors in the base, this additional electric field is negligibly small compared with that due to the acceptor ions. The fast increase in current with applied voltage ends when the additional electric field becomes dominant. The electrons injected into the base finally outnumber the acceptor ions. It is then that the current becomes fully space-charge-limited. The potential distribution in the base region for different applied voltages is shown in Fig. 8(b).

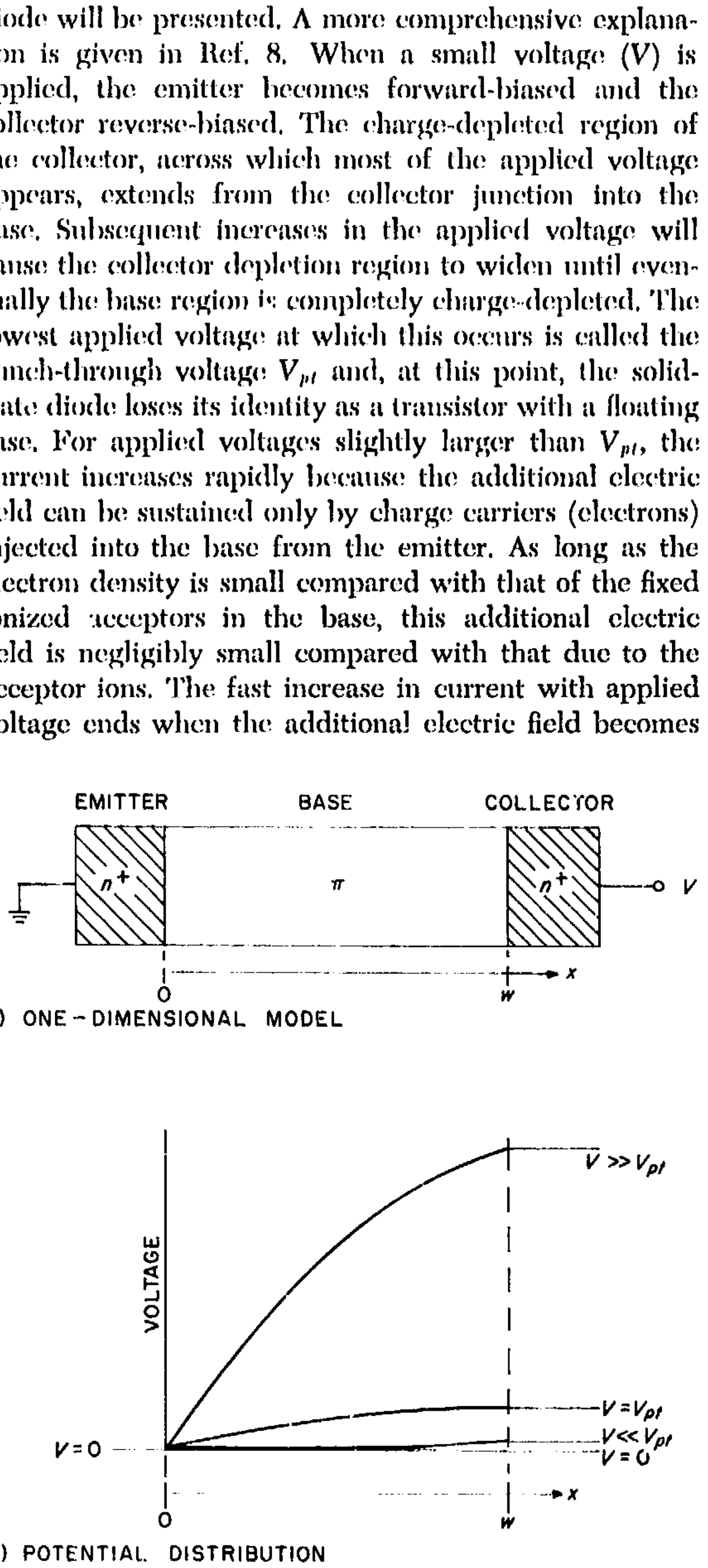


Fig. 8. An alloyed n^+-π-n^+ solid-state diode

c. Fabrication of solid-state diodes. Desirable features of an n^+-π-n^+ solid-state diode are that the $n^+\pi$ junctions be flat, parallel, abrupt and symmetric, and also that the lateral dimensions of the junction be large compared with the base width. This is accomplished by alloying two $n^+\pi$ junctions, one on each side of a [111]-oriented thin π-type wafer, such that they are of equal and sufficiently large area and directly opposite each other. A more detailed account of the procedures used in fabricating solid-state diodes is as follows: An ingot of π-type germanium oriented in the [111] direction is cut into slices of various thickness (10 to 20 mils) on a diamond saw. Each slice is further reduced to a specific thickness by lapping it mechanically in a stepwise manner with grinding compound of progressively smaller grit size, and is then lightly etched chemically with an etchant of 3 parts concentrated HNO_3 and 1 part concentrated HF. These slices are used to fabricate solid-state diodes of various base width. Wafers of 2.5 mm to the side, obtained by dicing the slices on a wire saw, are cleaned with trichloroethylene, acetone, distilled water, hydrofluoric acid, and hydrogen peroxide, respectively, and etched to a chemical polish with concentrated 3 HNO_3 : 1 HF etchant. In order to obtain accurately positioned junctions of well-defined area, a special masking technique was developed. The required geometry of the alloying substrate of the wafer is displayed by masking the other portions of the substrate with a thin film of SiO which is deposited in a vacuum of 2×10^{-6} torr; Fig. 9(a). This substrate is shielded during the deposition

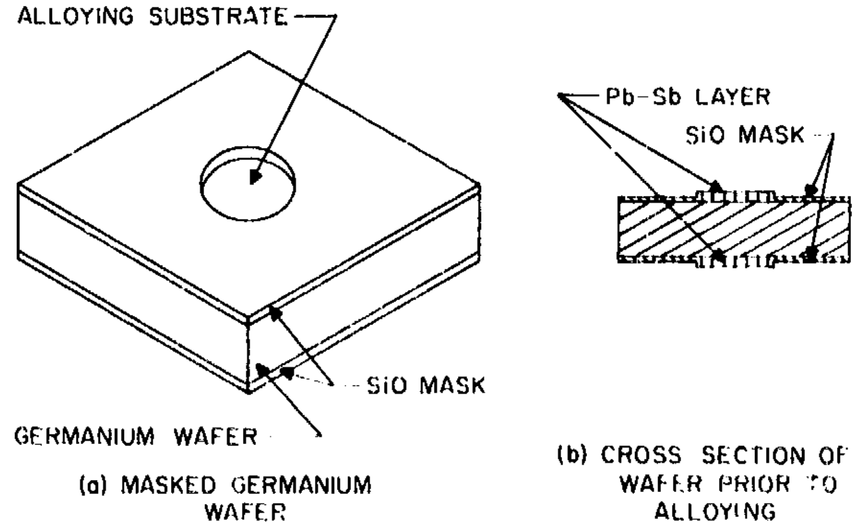


Fig. 9. Structure of a masked germanium wafer

by a temporary mask, which in this case is a 5 mil tantalum circular platelet of 0.7-mm diameter. The masked wafers are heat-cleaned at 750°C for 15 min in a vacuum of 10^{-6} torr and are then covered by a thin evaporated layer of 99.99% chemically pure antimony-lead mixture (10% Sb, 90% Pb, by weight) in order to prevent contamination to the heat-cleaned surfaces and thus insure good wetting during alloying. The metallic layer is removed everywhere except in the vicinity of the alloying substrate and the wafer as shown in Fig. 9(b) is ready for alloying.

Extra lead-antimony material is added to the metallic layer in pellet form. The alloying process which is a two-step operation takes place in a hydrogen atmosphere of a water-cooled horizontal strip furnace. First, the wafer is heated to 600°C such that the added pellet will have flowed to cover completely the alloying substrate. Second, the wafer is turned over and a similar procedure is followed except, now, top and bottom junctions are both alloyed at 675°C as shown in Fig. 10. The cooling rate during alloying is approximately constant and about 20°C/min. Subsequently the alloyed structure is cleaned for about 5 min in a solution containing approximately equal volumes of acetic acid and hydrogen peroxide to remove the lead-antimony at the periphery of the junction. The SiO mask is dissolved in hydrofluoric acid and the exposed surface is oxidized for 3 min in hydrogen peroxide just prior to the final etching which takes place for 5 sec in a concentrated 3 HNO_3 : 1 HF solution. Surface passivation is obtained by putting the wafer in hydrogen peroxide for 3 min. A cross-sectional representation of the fabricated structure is given in Fig. 11.

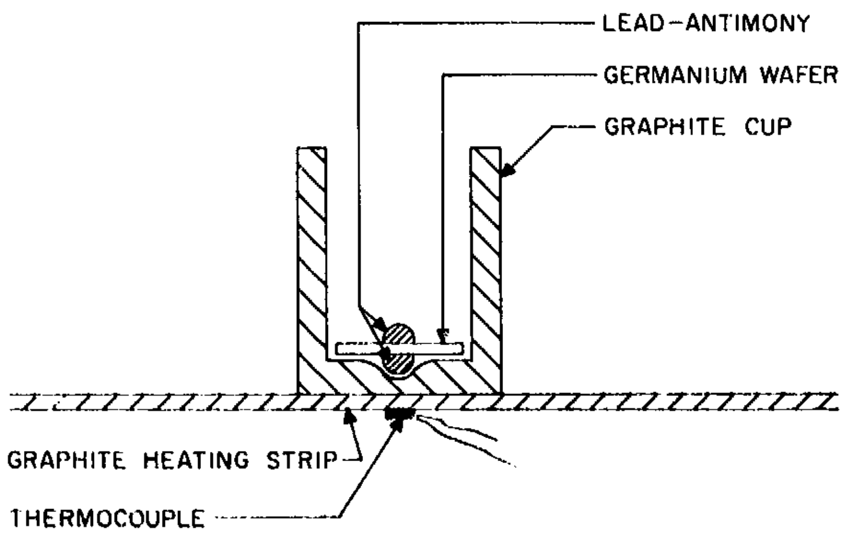


Fig. 10. Alloying of the n^+-π-n^+ solid-state diode in a graphite cup, cross-sectional view

Thermally-shocked 4-mil gold wires in contact with the emitter and collector cause localized melting of the lead-antimony mixture and thus make electrical bonds. These thermal shocks are produced by passing about a 150-amp surge current through the 4-mil wires for approximately 1 msec. The alloyed structure is encased in an air-tight holder. External emitter and collector contacts are made by soldering the gold wires with indium to the holder terminals. A liquid mixture of thallium, indium and gallium placed between the wafer and the holder makes the external electrical base connection.

It is important to be able to make n^+-π-n^+ structures of various base width and junction area. The former is accomplished by using wafers of various thicknesses and/or by alloying pellets of different size. The latter is accomplished by selecting platelets of various radii for the temporary masks.

Preliminary electrical measurements, such as the junction depletion capacitance, the forward and reverse biased junction *I-V* characteristics, and the punch-through voltage measurements, indicate that the fabricated solid-state diodes are adequate for making pure scl electron-current measurements. These results will be published in a future *Space Programs Summary*.

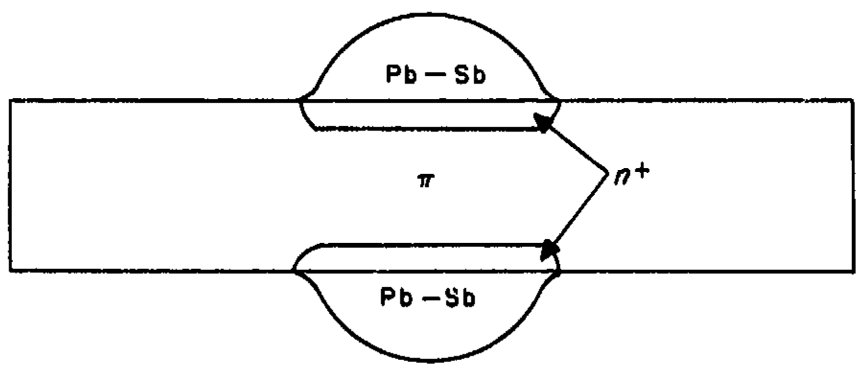


Fig. 11. Cross-sectional view of the alloyed n^+-π-n^+ solid-state diode

References

1. Crittenden, E. C., Hudimac, A. A., and Strough, I. R., *Review of Scientific Instruments, 22:* 972, 1951.
2. Oguey, H. J., *Review of Scientific Instruments, 31:* 701, 1960.
3. Chu, K., *IBM Research Report,* RC 617, January 1962.
4. Reimer, L., *Zeitschrift für Physik, 148; 527, 1957.*
5. Rose, A., and Smith, R. W., "Space-Charge-Limited Currents in Single Crystals of Cadmium Sulfide," *Physical Review, 97;* 1531–1537, 1955.
6. Dacey, G. C., "Space-Charge-Limited Hole Current in Germanium," *Physical Review, 90: 759–764,* 1953.
7. Wright, G. T., "Space-Charge-Limited Currents in Insulating Materials," *Nature, 182:* 1296–1297, 1958.
8. Shumka, A., *Space-Charge-Limited Current in Germanium,* Division of Engineering and Applied Science Report, California Institute of Technology, May 1964.

BLANK PAGE

ENGINEERING MECHANICS DIVISION

VII. Materials

A. Solid State Materials: Structure and Magnetic Properties of Glassy Carbon

D. B. Fischbach

In connection with investigations[1] of the high-temperature mechanical properties of glassy carbon[2], a brief study was made of the effect of heat treatment and deformation on this material's X-ray structure and magnetic properties. Two grades of glassy carbon—GC-20 and GC-30—were used which had been heat-treated by the manufacturer to 2000 and 3000°C, respectively. Samples from both the undeformed butt and deformed gage portions of tensile specimens which had been tested at temperatures up to 2900°C were examined at room temperature. Samples of each material were also measured after heat treatment at 3000 and 3200°C. X-ray diffraction data were obtained by the Debye-Scherrer technique, using Cu Kα radiation. The interlayer spacing, d, was determined from the (004) line—except for GC-20 samples tested below 2200°C where it was necessary to use the (002) line. Approximate values of the apparent crystallite layer diameter, L_a, were calculated from the displacement of the (11) peak, using a formula by Warren (Ref. 1). The magnetic susceptibility was measured by the Faraday method. Times during which the samples were at temperature varied from 20 to 60 min.

The results are shown in Fig. 1, where the open points refer to GC-20 and the solid points to GC-30. Deformed gage section samples are indicated by flags on the plotted points. The interlayer spacing of GC-20 decreased with heat treatment above 2200°C, reaching a value of 3.42 A after 20 min at 3200°C; however, the interlayer spacing of GC-30 was 3.42 A regardless of treatment. Thus, no appreciable layer stacking order developed in either material. L_a increased with heat treatment above 2700°C in GC-20 and above 3000°C in GC-30. In GC-20, tensile deformation $\gtrsim$ 10% elongation resulted in somewhat larger L_a values than heat treatment alone. In GC-30, L_a was not affected by deformation, which was small.

[1]*SPS 37-24*, Vol. IV, p. 46; *SPS 37-25*, Vol. IV, p. 43; *SPS 37-26*, Vol. IV, p. 70; *SPS 37-27*, Vol. IV, pp. 32–33.

[2]Tokai Electrode Manufacturing Co., Nagoya, Japan.

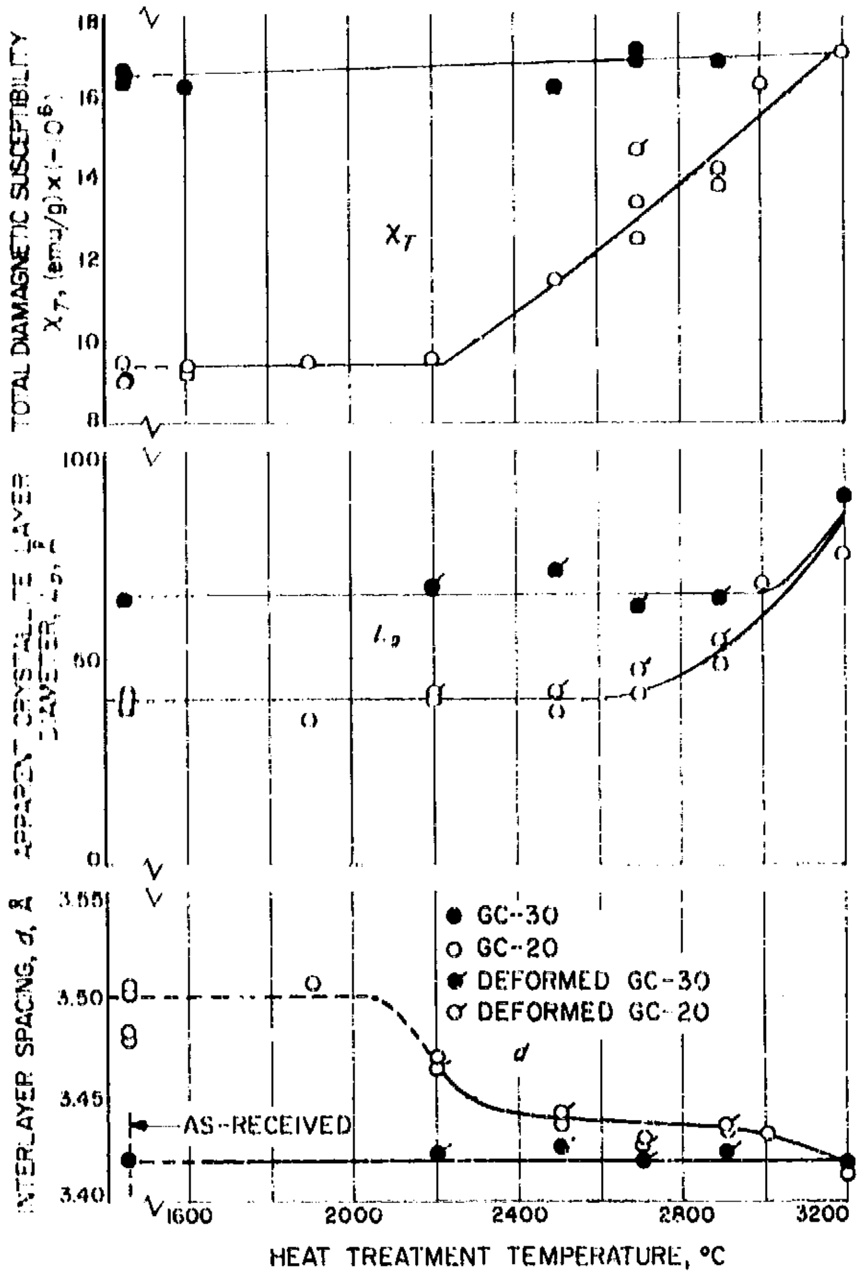


Fig. 1. Dependence of total diamagnetic susceptibility, interlayer spacing, and apparent crystallite layer diameter of glassy carbons GC-20 and GC-30 on heat-treatment (and deformation) temperature

The total diamagnetic susceptibility (sum of the susceptibilities measured in three perpendicular directions) of GC-20 increased with heat treatment above 2200°C, reaching values characteristic of GC-30 after treatment at 3000 to 3200°C. As expected, the susceptibility of GC-30 was not appreciably affected by heat treatment in this range. For GC-20, heavy tensile deformation appeared to produce larger susceptibilities than heat treatment alone. As-received and heated-only glassy carbons were magnetically isotropic. However, deformed samples were anisotropic, exhibiting anisotropy ratios (maximum over minimum susceptibility) as large as 1.8 after 33% elongation, with the minimum susceptibility parallel to the tensile axis. This indicates that some alignment of the layer planes parallel to the tensile axis is produced by the deformation.

The magnetic susceptibility was independent of field strength for glassy carbons heat-treated at 2700°C or higher, but as-received GC-20 showed some field-strength dependence. In most cases this was slight, but one lot was paramagnetic at the usual measurement field strength (15.5 kilogauss). However, the susceptibility extrapolated at infinite field strength to a diamagnetic value in agreement with that obtained for other lots of GC-20. Such field-strength dependence is characteristic of ferromagnetic impurities. Spectrochemical analysis indicated traces of Fe in as-received GC-20 but not in GC-30, which is in agreement with the magnetic results.

The values of the X-ray structure parameters and their variation with heat treatment are in general agreement with recently published data[3] (Ref. 2). These values show that glassy carbon is hard and non-graphitizing. Like other carbons, the diamagnetism of glassy carbons appears to be an increasing function of L_a. However, the behavior of glassy carbons differs somewhat from that of most soft (graphitizing) carbons. For heat treatment below 2700°C, the susceptibility of GC-20 increased without change in L_a. (This may result from evaporation of para- and ferromagnetic impurities, or from disappearance of carrier trapping centers.) In addition, the susceptibility attained higher values for $L_a < 100A$ than are attained in most soft carbons in this range. In the latter respect, the behavior of the glassy carbons is very similar to that of cellulose carbons (Ref. 4), and may be characteristic of hard carbons.

[3]"Structure of Glassy Carbon," paper by T. Noda and M. Inagaki presented at the Symposium on Carbon, Tokyo, July 1964.

References

1. Warren, B. E., "X-Ray Diffraction in Random Layer Lattices," *The Physical Review*, Vol. 59, p. 693, 1941.

2. Yamada, S., Sato, H., Ishii, T., "Eigenschaften und Verwendung von Glasartigem Kohlenstoff," *Carbons*, Vol. 2, p. 253, 1964.

3. Adamson, A. F., Blayden, H. E., "Magnetic Studies of Graphite and Mesomorphous Carbons," *Industrial Carbon and Graphite Papers Conference, London, 1957*, p. 147, Pergamon Press, New York, 1958.

VIII. Lunar Spacecraft Development

A. High Impact Technology

J. L. Adams

As reported in *SPS 37-31*, Vol. IV, pp. 118–121, several tasks are underway in the JPL high impact program. The following article cites the accomplishments in each task during this reporting period (February 1 to March 31).

Ruggedized gas chromatograph. Table 1 shows component tests that were conducted. The chief component problems at this time are the dynamic capacitor and the pressure regulator and valves. The circuit design, the prototype mechanical design, and the preliminary packaging layouts are completed.

Ruggedized S-band transmitter. The prototype front-end of the transmitter was impact tested at 3400 *g* from 110 ft/sec in all three principle directions. During this series of tests, the high-impact crystal failed because of an inadequate case solder joint. The front-end was then impacted at 9500 *g* from 190 ft/sec in both axial directions. Except for the crystal, the unit operated satisfactorily after all impacts. The second stage of the transmitter, which contains amplification and multiplication up to a frequency in the VHF band, is ready for impact testing.

Batteries. A small prototype silver-zinc cell was constructed from plates and separator material taken from a silver-zinc battery. The case of the prototype battery was designed to prohibit movement of the plate stack. The inside walls were grooved to allow the necessary circulation. This cell, in a partially uncharged condition, survived an impact of 4200 *g* from 110 ft/sec normal to the plates, with no apparent degradation; however, later tests in all three principal directions at 7500 *g* from 180 ft/sec resulted in destruction of the battery, with severe mechanical damage to the plates.

Mechanism. A rebuilt Gaylord-Rives size 8 synchronous motor survived the equivalent of 10,000 *g* from 200 ft/sec. This motor utilizes spring-centered bearings to shift the impact loads from the bearings to both the

Table 1. Test results of ruggedized gas chromatograph components

Component	Sample size	Impact level	Impact direction	Comments
Texas Instrument ⅛-, ¼-, and ½-w 1% glass-carbon resistors	18	10,000 *g* 190 ft/sec	Axially and transversely	Survived
Kinelogic dynamic capacitor	1	5500 *g* 120 ft/sec	Axially and transversely	Performance slightly degraded after each test. Corrected to some extent by overdriving
		10,000 *g* 200 ft/sec	Axially and transversely	Axial impact seriously degraded capacitor performance. Transverse impact destroyed capacitor
GC ionization detector	1	10,000 *g* 190 ft/sec	Axially	Survived
Victorene High-Meg resistors	15	10,000 *g* 190 ft/sec	Axially and transversely	Survived
GC Leonard pressure regulator	1	3500 *g* 115 ft/sec	Axially and transversely	Survived
		5600 *g* 115 ft/sec	Axially	Leakage increased beyond desirable value
Sprague 150-B solid-slug B size tantalum capacitors	20	10,000 *g* 190 ft/sec	Axially and transversely	Survived

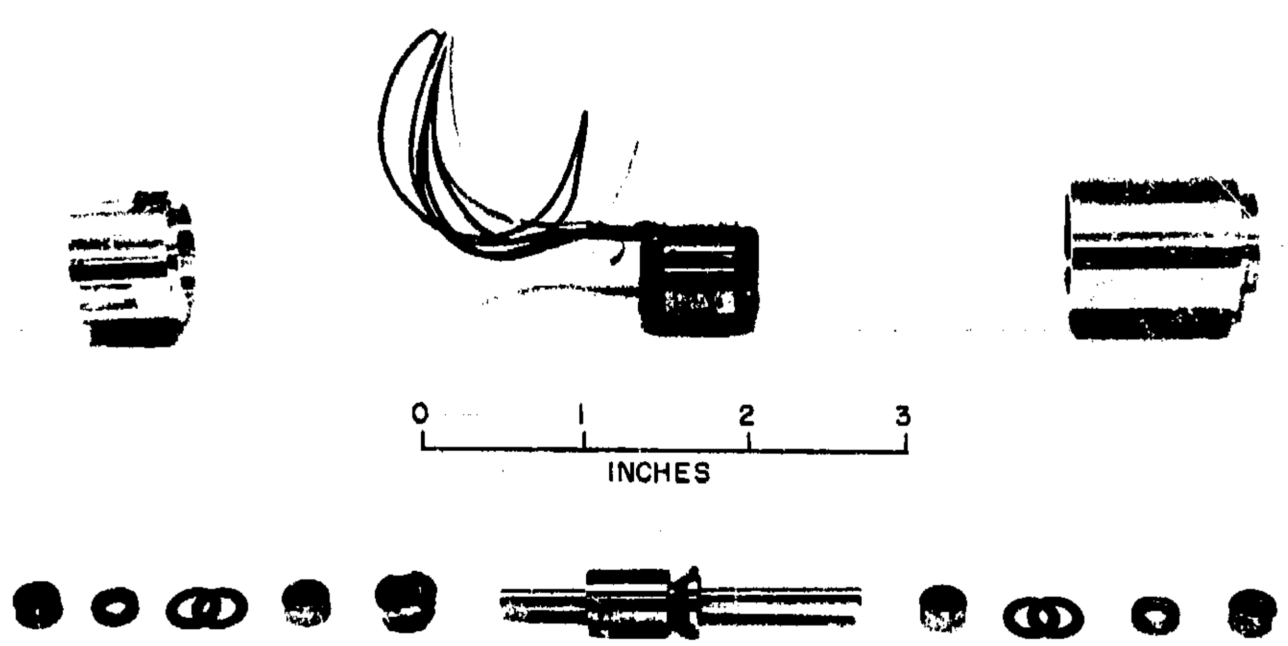


Fig. 1. Components of rebuilt Gaylord-Rives Size 8 synchronous motor

armature and case. Fig. 1 shows the motor's components. Development is underway on a prototype belt-driven tape recorder transport capable of surviving 10,000 *g* from 500 ft/sec. This recorder will utilize a larger version of the motor shown in Fig. 1.

High impact facility. The 6-in. D compressed-air gun is operational and is currently being used for investigating the energy absorbing capabilities of balsa wood at 500 ft/sec. A compressed-air gun with a 22-in. D bore is being developed which will operate identically to the

6-in. gun except that the former will have a single barrel mounted upon a carriage designed to absorb the recoil. The gun will be used for two purposes:

(1) To accelerate projectiles containing test specimens to 500 ft/sec for subsequent deceleration against a crushable material. This operation is identical to that used for the 6-in. gun.

(2) To accelerate a carriage which will be stopped before the test specimen is allowed to impact. With this carriage, packages up to 5 ft D can be tested. This lack of size constraint, coupled with the large test-weight capability of the gun (approx. 500 lb to 500 ft/sec, including the carriage or projectile) will enable testing of large pieces of capsule equipment, prototype payload assemblies, and large segments of energy absorbing structure.

IX. Applied Mechanics

A. Radiation-Conduction Coupled Heat Transfer Investigations

J. A. Plamondon

For some time there has been concern over the accuracy of analytical prediction of radiant heat exchange for complex geometries such as those found in spacecraft. This concern has resulted from differences between (1) the analytical predictions, (2) data obtained experimentally in space chambers, and (3) data obtained from flight. A study has been undertaken to explain and eliminate some of the causes for these differences. For this study, a simple geometry was chosen for investigation to obtain agreement between analytical and experimental methods, and develop insight into the causes of disagreement.

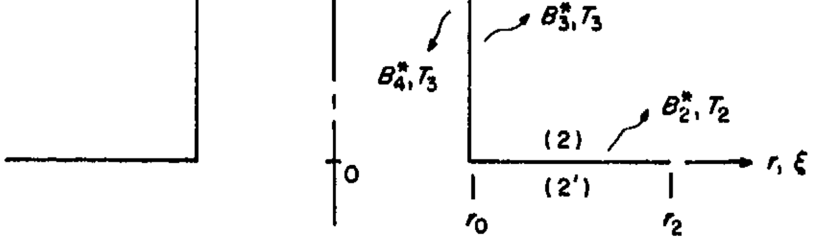


Fig. 1. Section of circumferentially-finned cylinder

The geometry being investigated (Fig. 1) is an axially-symmetric finned cylinder. It was chosen because of its similarity to the *Ranger* TV tower for which both flight and experimental chamber data disagreed with analytical prediction. This choice of geometry was suitable because it contains those characteristics of radiant heat exchange found in complex spacecraft hardware.

To date, work performed on this study has been primarily analytical and preparatory. Thus far, analytical work has been accomplished for two cases of radiant heat exchange where: (1) temperature, radiosity, and solar heat load are assumed uniform on each surface, and (2) radiosity and temperature distributions on each surface are accounted for. Cases (1) and (2) were programmed for the IBM 7094 computer and solved. Conduction coupling has been added to Case (2) and is being programmed for numerical solution. No experimental work has been accomplished, due to nonavailability of an appropriate small space chamber; however, such a facility is being assembled.

Analysis has proceeded from the simple uniform temperature and radiosity situation to the complex radiation-conduction coupled, non-uniform situation. In Case (1) temperature, radiosity, and solar irradiance were assumed to be uniform for each surface of the circumferentially-finned cylinder. In addition, the first analysis was formulated within the framework of the diffuse surface idealization. In this analysis, solar heating was assumed to be introduced along the axis of symmetry, and no temperature gradient was assumed to exist through the walls. The thermal state of each surface was computed on the basis of thermodynamic equilibrium with its environment, which was assumed to consist of the external solar heat load, space at absolute zero degrees, and the adjacent surfaces. This particular analysis was undertaken (in addition to previously mentioned reasons) for sizing the geometry and choosing surface properties such that thermodynamic equilibrium occurs near room temperature for an external solar input of approximately 1 solar constant. Since the derivation of the governing equations is beyond the scope of this summary, only the following final governing equations are presented:

$$\left(1-\frac{\epsilon_1^2}{\epsilon_1+\epsilon_1'-\epsilon_1\epsilon_1'}\right)B_1^*-\rho_1F_{1-2}B_2^*-\rho_1F_{1-3}B_3^*=\frac{\epsilon_1(1-\epsilon_1)\alpha_{1s}^1}{\epsilon_1+\epsilon_1'-\epsilon_1\epsilon_1'}Q_1^* \tag{1}$$

$$-\rho_2F_{2-1}B_1^*+\left(1-\frac{\epsilon_2^2}{\epsilon_2+\epsilon_2'-\epsilon_2\epsilon_2'}\right)B_2^*-\rho_2F_{2-3}B_3^*=\left(\rho_{2s}+\frac{\epsilon_2(\alpha_{2s}-\epsilon_2)}{\epsilon_2+\epsilon_2'-\epsilon_2\epsilon_2'}\right)Q_2^* \tag{2}$$

$$-\rho_3F_{3-1}B_1^*-\rho_3F_{3-2}B_2^*+\left(1-\frac{\epsilon_3^2(1-\epsilon_4)}{\epsilon_3+\epsilon_4-2\epsilon_3\epsilon_4}\right)B_3^*-\frac{\epsilon_3\epsilon_4(1-\epsilon_3)}{\epsilon_3+\epsilon_4-2\epsilon_3\epsilon_4}B_4^*=\frac{\epsilon_3(1-\epsilon_3)(\alpha_{4s}-\epsilon_4)}{\epsilon_3+\epsilon_4-2\epsilon_3\epsilon_4}Q_4^* \tag{3}$$

$$-\frac{\epsilon_3\epsilon_4(1-\epsilon_4)}{\epsilon_3+\epsilon_4-2\epsilon_3\epsilon_4}B_3^*+\left(1-\rho_4F_{4-4'}-\frac{\epsilon_4^2(1-\epsilon_3)}{\epsilon_3+\epsilon_4-2\epsilon_3\epsilon_4}\right)B_4^*=\left(\rho_{4s}+\frac{\epsilon_4(1-\epsilon_3)(\alpha_{4s}-\epsilon_4)}{\epsilon_3+\epsilon_4-2\epsilon_3\epsilon_4}\right)Q_4^* \tag{4}$$

where B^* is the dimensionless radiosity; F_{i-j} is the view factor between surfaces i and j; Q^* is the dimensionless solar heat load (solar irradiance); ϵ, α, and ρ are the properties of the surfaces; subscripts 1, 2, 3, and 4 refer to the surfaces as shown in Fig. 1; and subscript s denotes property to solar spectrum. Dimensionless temperature *to the fourth power* is related to Eqs. (1) to (4) by the following:

$$T_1^*=\frac{\epsilon_1B_1^*+(1-\epsilon_1)\alpha_{1s}'Q_1^*}{\epsilon_1+\epsilon_1'-\epsilon_1\epsilon_1'} \tag{5}$$

$$T_2^*=\frac{\epsilon_2B_2^*+(\alpha_{2s}-\epsilon_2)Q_2^*}{\epsilon_2+\epsilon_2'-\epsilon_2\epsilon_2'} \tag{6}$$

$$T_3^*=T_4^*=\frac{\epsilon_3(1-\epsilon_4)B_3^*+\epsilon_4(1-\epsilon_3)B_4^*+(1-\epsilon_3)(\alpha_{4s}-\epsilon_4)Q_4^*}{\epsilon_3+\epsilon_4-2\epsilon_3\epsilon_4} \tag{7}$$

Dimensional temperatures can be obtained from

$$T(°\mathrm{R})=\left(\frac{T^*Q_0}{\sigma}\right)^{1/4} \tag{8}$$

and dimensional radiosities from

$$B(\mathrm{Btu/hr\ ft^2})=B^*Q_0 \tag{9}$$

where Q_0 is the solar heat load in Btu/hr ft^2 and σ is the Stefan-Boltzmann constant. These equations were solved numerically and resulted in a configuration with the dimensionless size and surface properties as shown in Table 1.

Table 1. Dimensionless size and surface properties of circumferentially-finned cylinder

Size	Surface properties	
$N = l/r_0$ 4.000	Surface 1′ black	$\epsilon_1' = 0.9400$ $\alpha_{1s}' = 0.9600$
$\eta_0 = r_1/r_0$ 2.000	Surface 1 white	$\epsilon_1 = 0.8500$
$\xi_0 = r_2/r_0$ 2.646	Surface 2′ gold	$\epsilon_2' = 0.0450$
$Q_1^* = Q_1/Q_0$ 1.000	Surface 2 black	$\epsilon_2 = 0.9400$ $\alpha_{2s} = 0.9600$
$Q_2^* = Q_2/Q_0$ 0.500	Surface 3 polished aluminum	$\epsilon_3 = 0.0400$
$Q_4^* = Q_4/Q_0$ 0.100 (= to 7½-deg decollination angle)	Surface 4 black	$\epsilon_4 = 0.9400$ $\alpha_{4s} = 0.9600$

With these equations solved, Case (2) was undertaken to determine the temperature distribution along each surface and to see how this distribution differs from the bulk-type temperature computed in Case (1). Again the same assumptions were made as in Case (1), except that the uniform temperature and radiosity restriction was removed. The second analysis resulted in a set of seven simultaneous equations: four integral equations in terms of temperature and radiosity, and three algebraic equations relating temperature and radiosity. Again the derivation of these equations is beyond the scope of this summary. The final dimensionless governing equations are:

$$B_1^*(\eta) = \epsilon_1 T_1^*(\eta) + \rho_1 \int_1^{\xi_0} \xi\, B_2^*(\xi)\, K(\eta,\xi)\, d\xi + \rho_1 \int_0^N B_3^*(\beta)\, K(\eta,\beta)\, d\beta \tag{10}$$

$$B_2^*(\xi) = \epsilon_2 T_2^*(\xi) + \rho_2 \int_1^{\eta_0} \eta\, B_1^*(\eta)\, K(\eta,\xi)\, d\eta + \rho_2 \int_0^N B_3^*(\beta)\, K(\xi,\beta)\, d\beta + \rho_{2s} Q_2^*(\xi) \tag{11}$$

$$B_3^*(\beta) = \epsilon_3 T_3^*(\beta) + \rho_3 \int_1^{\eta_0} \eta\, B_1^*(\eta)\, K(\eta,\beta)\, d\eta + \rho_3 \int_1^{\xi_0} \xi\, B_2^*(\xi)\, K(\xi,\beta)\, d\xi \tag{12}$$

$$B_4^*(\beta) = \epsilon_4 T_3^*(\beta) + \rho_4 \int_0^N B_4^*(\gamma)\, K(\gamma,\beta)\, d\gamma + \rho_{4s} Q_4(\beta) \tag{13}$$

$$T_1^*(\eta) = \frac{\epsilon_1 B_1^*(\eta) + (1-\epsilon_1)\,\alpha'_{1s}\, Q_1^*(\eta)}{\epsilon_1 + \epsilon'_1 - \epsilon_1 \epsilon'_1} \tag{14}$$

$$T_2^*(\xi) = \frac{\epsilon_2 B_2^*(\xi) + (\alpha_{2s} - \epsilon_2)\, Q_2^*(\xi)}{\epsilon_2 + \epsilon'_2 - \epsilon_2 \epsilon'_2} \tag{15}$$

$$T_3^*(\beta) = \frac{\epsilon_3 (1-\epsilon_4)\, B_3^*(\beta) + \epsilon_4 (1-\epsilon_3)\, B_4^*(\beta)}{\epsilon_3 + \epsilon_4 - 2\epsilon_3\epsilon_4} + \frac{(1-\epsilon_3)(\alpha_{4s} - \epsilon_4)\, Q_4^*(\beta)}{\epsilon_3 + \epsilon_4 - 2\epsilon_3\epsilon_4} \tag{16}$$

The nomenclature is the same as before; however, the view factors, F, have been replaced by the kernels, K, which denote the view functions between incremental areas. Also, the B^* and T^* are now functions of the coordinates of the individual surfaces. The expressions for the kernels will not be presented because of their length. These equations were solved numerically using a Simpson's Rule and a simple iteration procedure in which the integrals were divided into n finite areas.

Figs. 2, 3, and 4 show the temperature and radiosity distributions on the upper fin, lower fin, and cylinder. The horizontal lines are the temperatures and radiosities computed from Case (1); the curved and discontinuous lines are from Case (2). Except for the upper fin, significant differences exist between the predictions of the two Cases. Figs. 5, 6, and 7 show the effect on the results obtained

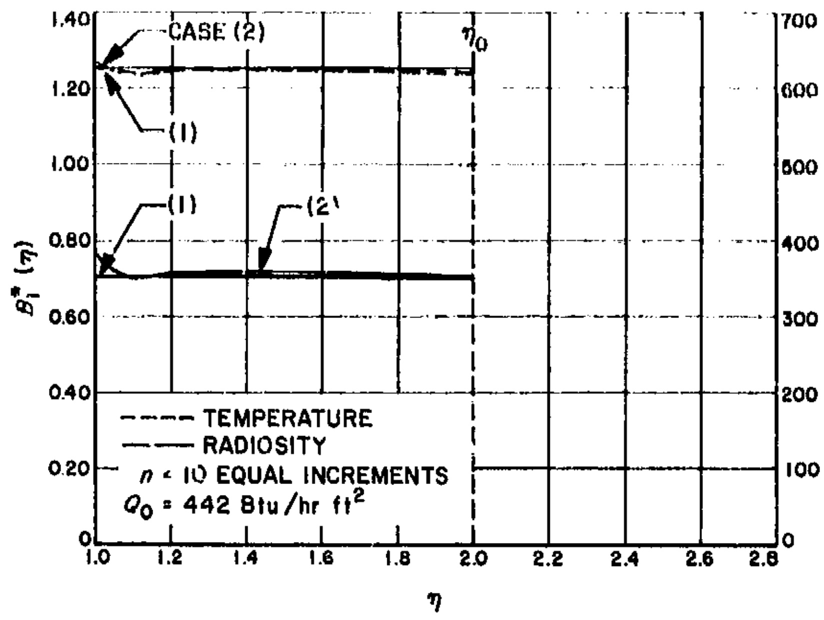


Fig. 2. Temperature and dimensionless radiosity distribution, upper fin

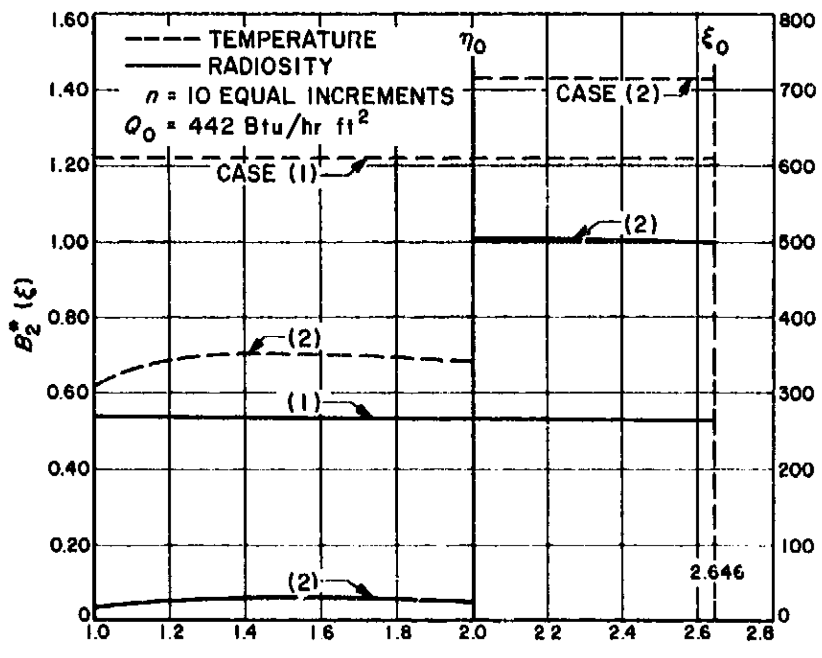


Fig. 3. Temperature and dimensionless radiosity distribution, lower fin

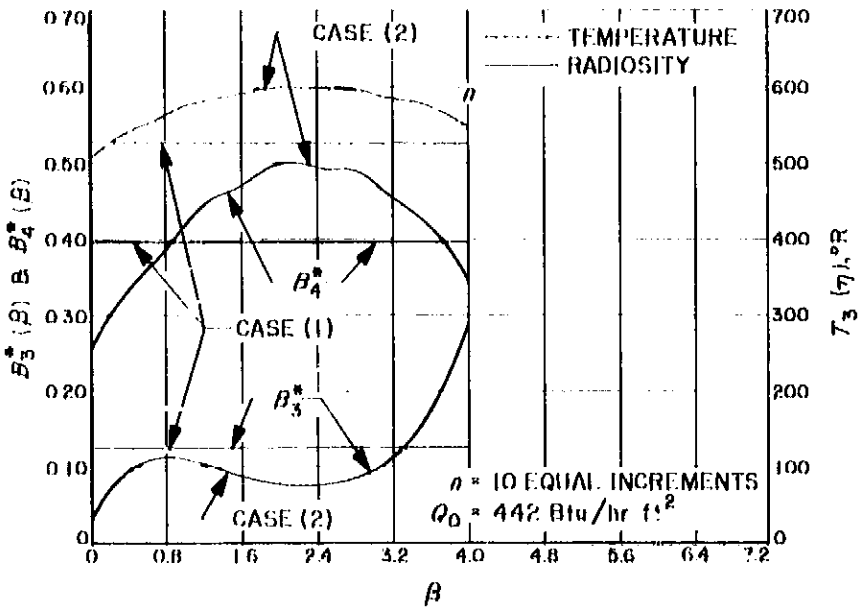


Fig. 4. Temperature and dimensionless radiosity distribution, cylinder

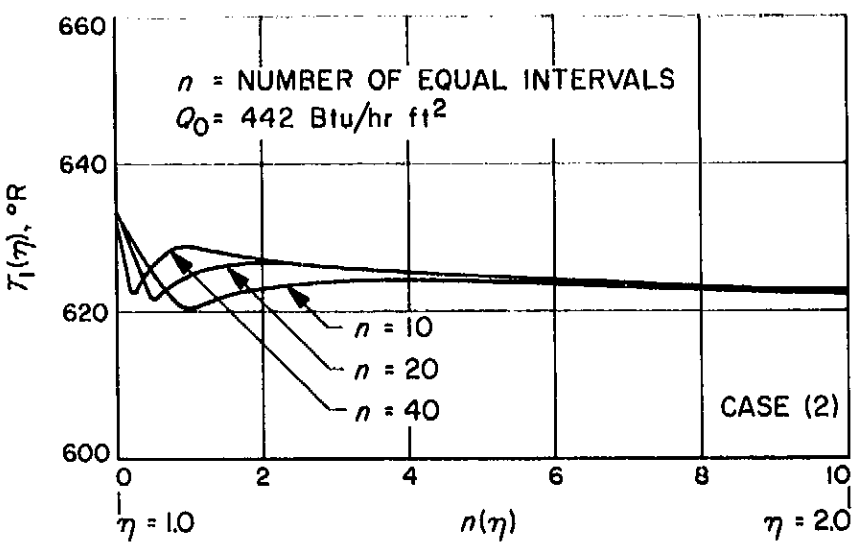


Fig. 5. Temperature distribution in the upper fin for various values of the integration interval, *n*

of the number of intervals into which the integrals in the integral equations are divided. The curves shown are for Case (2) with $n = 10$, 20, and 40. For n greater than 40, no differences were noted; consequently, 40 intervals seem sufficient for these calculations.

In Figs. 3 and 6, a discontinuity exists at $\xi = 2.0$ that is attributed to a discontinuity in the direct solar input as a result of shadowing by the upper fin. In the first analysis (uniform temperature and radiosity), the solar flux on the lower fin was incident uniformly across the entire fin; in the second analysis, the solar flux impinged only on that area of the lower fin exposed to direct sunlight. This partially accounts for the extremely large differences between the two analyses of temperature and radiosity observed on the lower fin.

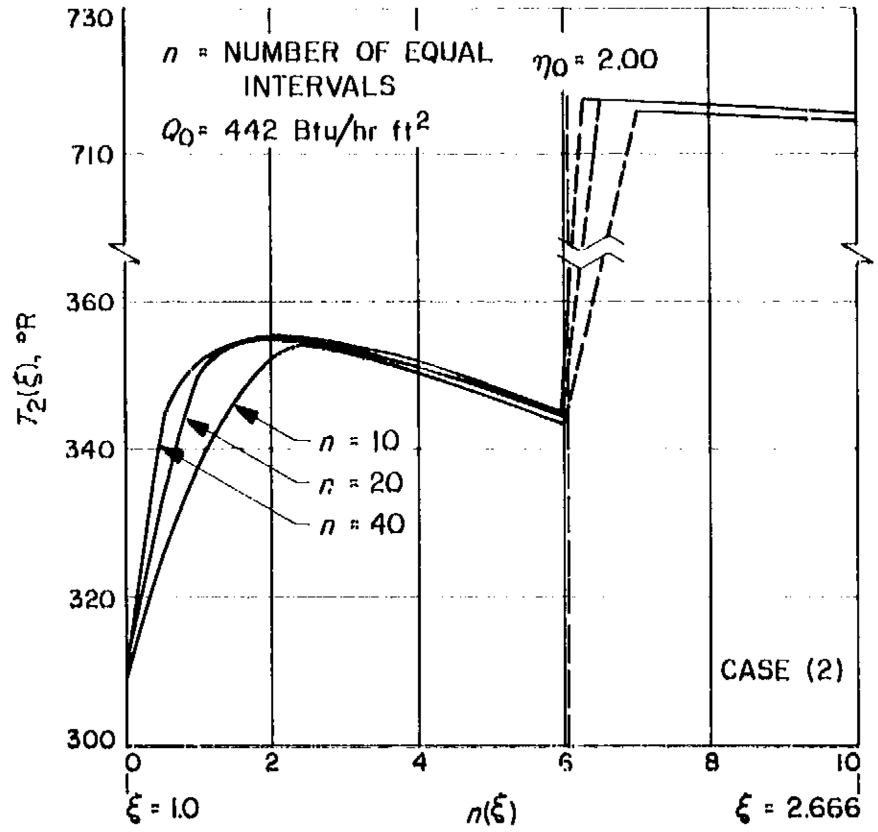


Fig. 6. Temperature distribution in lower fin for various values of the integration interval, *n*

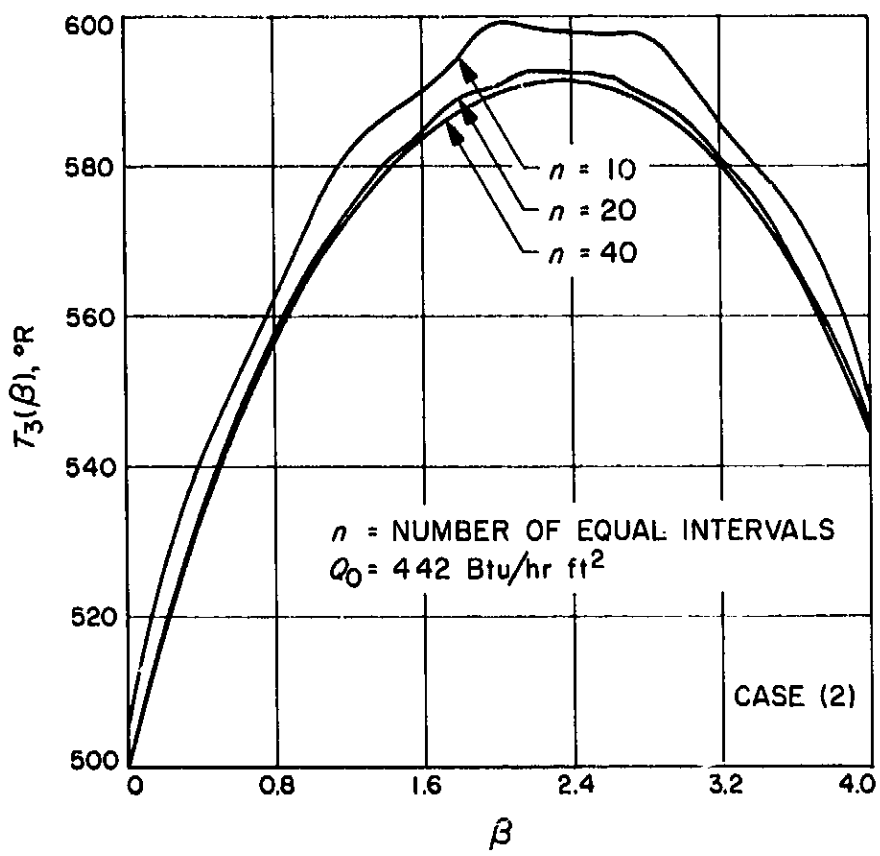


Fig. 7. Temperature distribution in cylinder wall for various values of the integration interval, *n*

Since large differences exist between the two analyses, a third analysis has been undertaken in which conduction coupling is included. Since conduction would tend to decrease the large temperature gradients that exist on both cylinder and fin in the second analysis, the results of the third analysis should lie somewhere between the results of the first two. It can be concluded that the results of either analysis could not, without further justification, be reliably used as a means for analytical prediction because of the large differences that exist.

B. Brazed Pyrolytic Graphite to Metal Joints for Advanced Liquid Propulsion Systems

C. E. Johnson

Pyrolytic graphite (PG) 100-lb thrust chambers of the type shown in Fig. 8 have been designed, fabricated, and successfully tested. To date, however, any uncooled connection of the PG chamber to a metal injector has been accomplished only by using an adapter ring made of ATJ graphite; the ring itself is very heavy, the cause attributed to large differential thermal expansions between the metal injector and PG. In an effort to optimize the injector/thrust chamber design, methods of brazing metal to pyrolytic graphite are being studied.

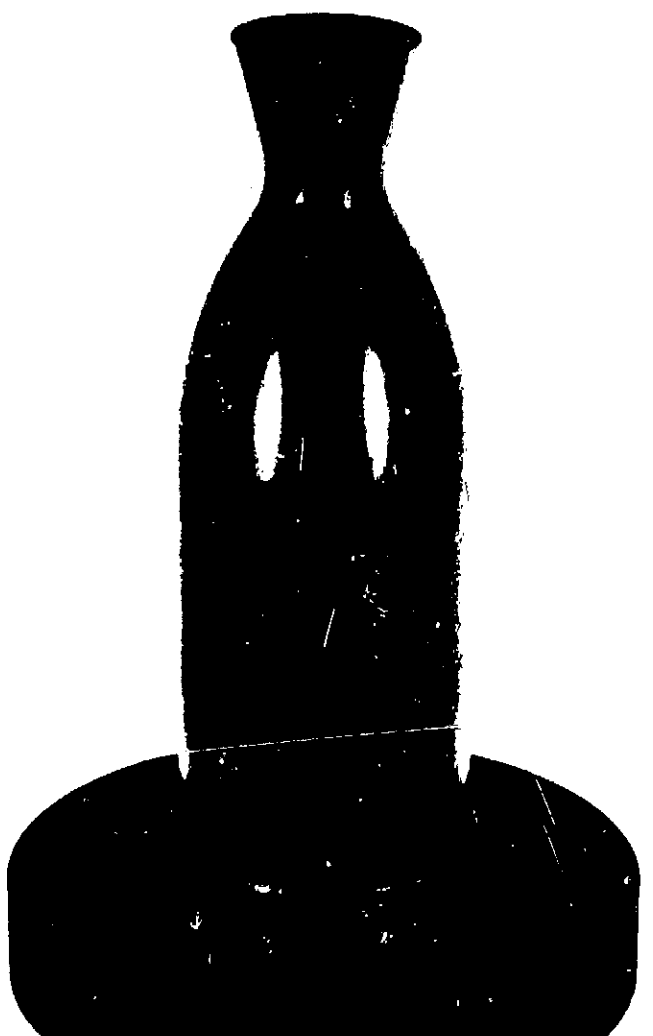


Fig. 8. 100-lb pyrolytic graphite thrust chamber with an ATJ graphite adapter ring

Braze materials (*Si*-2*Ti*, *Ti*–8.5*Si*, etc.) have been developed that will withstand burn temperatures of 2500°F and still retain structural strength. When metal and PG are brazed together, the oven temperature is at 2940°F; therefore, the structural criterion for designing the joint is the temperature differential between 2940°F and the lowest joint temperature experienced during flight, which is approximately 60°F.

Preliminary analytical investigations were conducted to show the feasibility of using a particular metal for the joint geometry shown in Fig. 9. A planar hoop-stress analysis was performed that yielded an approximate relation (Eq. 1) for determining feasibility and rank comparison for any metal/PG optimum joint design. This equation must be satisfied for joint feasibility to exist.

$$\frac{\sigma_M}{E_M} + \frac{\sigma_{PG}}{E_{PG}} \geqq \alpha_M - \alpha_{PG} - \frac{\alpha_c}{2}\frac{t_{PG}}{R} \qquad (1)$$

where

σ_M = tension yield stress in the metal at room temperature, psi

σ_{PG} = compression yield stress (along the grain) in the PG at room temperature, psi

E_M = elastic modulus of the metal at room temperature, psi

E_{PG} = elastic modulus (along the grain) of the PG at room temperature, psi

α_M = equivalent coefficient of thermal expansion of metal for a given temperature gradient, in./in.

α_{PG} = equivalent coefficient of thermal expansion of PG for a given temperature gradient along perimeter of cylinder (along the grain), in./in.

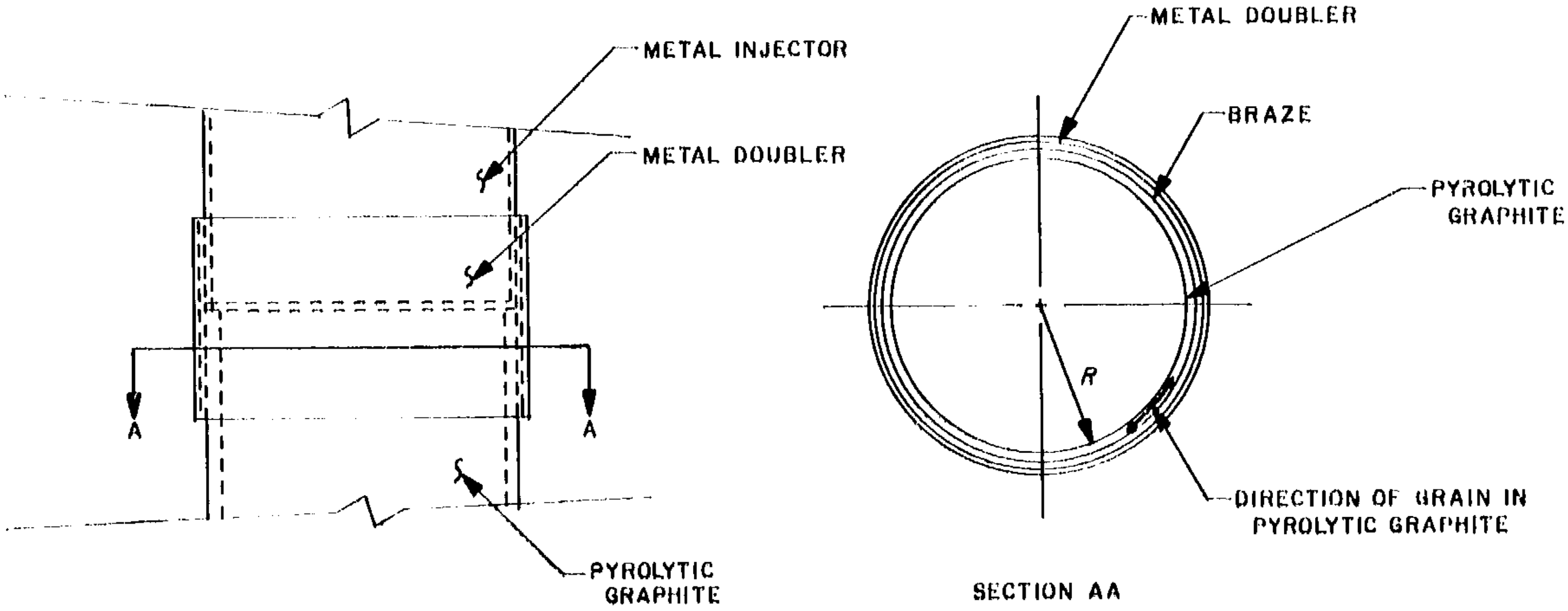


Fig. 9. Metal-pyrolytic graphite joint

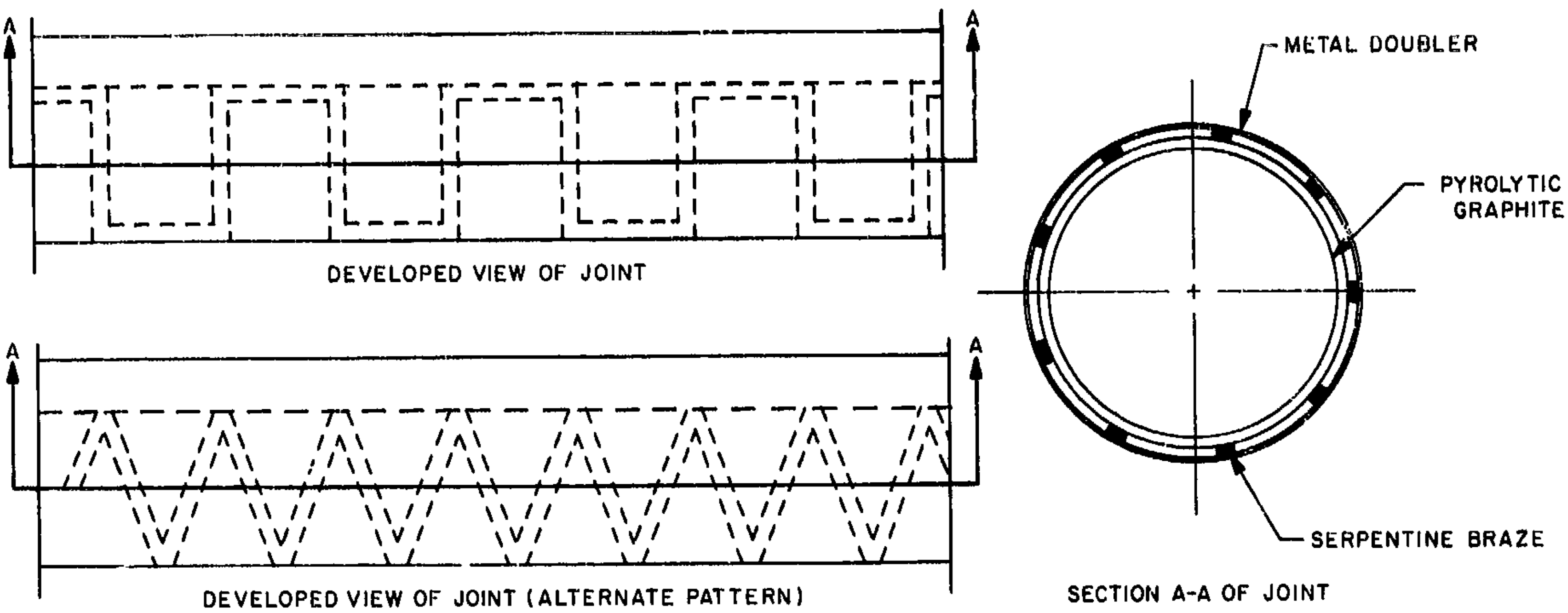


Fig. 10. Metal-pyrolytic graphite serpentine-braze joint

α_e = equivalent coefficient of thermal expansion of PG for a given temperature gradient in radial direction (cross the grain), in./in.

t_{PG} = thickness of PG, in.

R = mean radius of thin-walled brazed joint, in.

A relation similar to Eq. (1) can be written for preload in the joint at operating temperature:

$$\frac{\sigma_M}{E_M} + \frac{\sigma_{PG}}{E_{PG}} - \left(\frac{\sigma_{MT}}{E_{MT}} + \frac{\sigma_{PGT}}{E_{PGT}}\right) \geq \alpha_M - \alpha_{PG} - \frac{\alpha_e}{2}\frac{t_{PG}}{R} \quad (2)$$

where the subscript T denotes structural properties at temperature.

In comparing Eqs. (1) and (2), it is seen that the use of a preload at operating temperature has a detrimental effect on the structural adequacy of the joint. Good practice in the fabrication of closed shapes with PG dictates that the t/R ratio does not vary much from $t/R = 0.05$, due to inherent built-in stresses. Therefore, by inserting this value into Eqs. (1) and (2), decisions can be made for any size joint.

As a result of an investigation of the following refractory metals, it was determined that none could be used without general yielding occurring in the metal or PG.

(1) Tantalum.

(2) Tantalum-10 Tungsten.

(3) Tantalum-8 Tungsten-2.4 Hafnium.

(4) Molybdenum.

(5) Molybdenum-0.5 Titanium.

(6) Columbium.

(7) Columbium-10 Titanium-10 Molybdenum.

(8) Columbium-15 Tungsten-5 Molybdenum-1 Zirconium.

Based on this preliminary investigation, it was decided that to successfully braze metal to PG, the joint should be designed to relieve the excessive hoop stress involved. Also, the design should incorporate a seal to prevent leakage of thrust pressure.

Fig. 10 illustrates a metal/PG serpentine-braze joint designed to relieve excessive hoop stress caused by differential thermal expansion. The proposed design creates a flexible composite structure by reducing the amount of braze material between the metal and PG. Deformation, due to differential thermal contraction, can now be absorbed primarily by local bending in the metal instead of only membrane loads in the metal and PG. The internal pressure, between the metal and PG, is significantly reduced, thereby decreasing the general hoop-stress condition.

The serpentine-braze joint was roughly analyzed to determine the PG loading condition and whether or not the metal yields when deforming into the provided space. The mathematical model used is shown in Fig. 11. Curves of maximum stress (psi) and hoop load (lb.) vs. metal thickness (in.) were derived for any metal, temperature differential, and general braze pattern (Figs. 12 and 13). E is the elastic modulus of the metal at room temperature (psi), and α is the equivalent metal coefficient of thermal expansion (in./in.) for a given temperature gradient. Based on this analysis, it was determined that for the 100-lb thrust chamber (1) the metal will yield locally at the braze joints, and (2) the induced hoop and longitudinal stress in the PG is an order of magnitude less than allowable.

The serpentine-braze joint was also analyzed to determine whether or not the metal deforms into the space

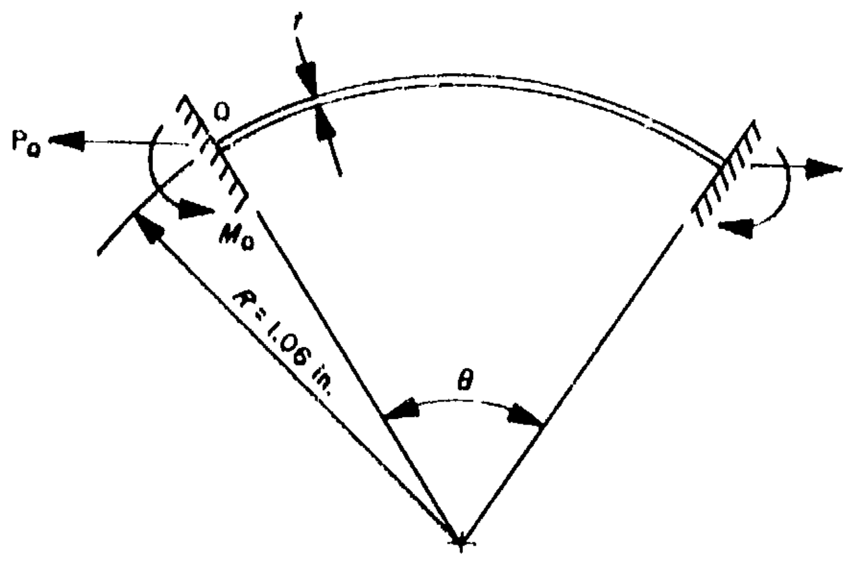


Fig. 11. A mathematical model, serpentine-braze joint

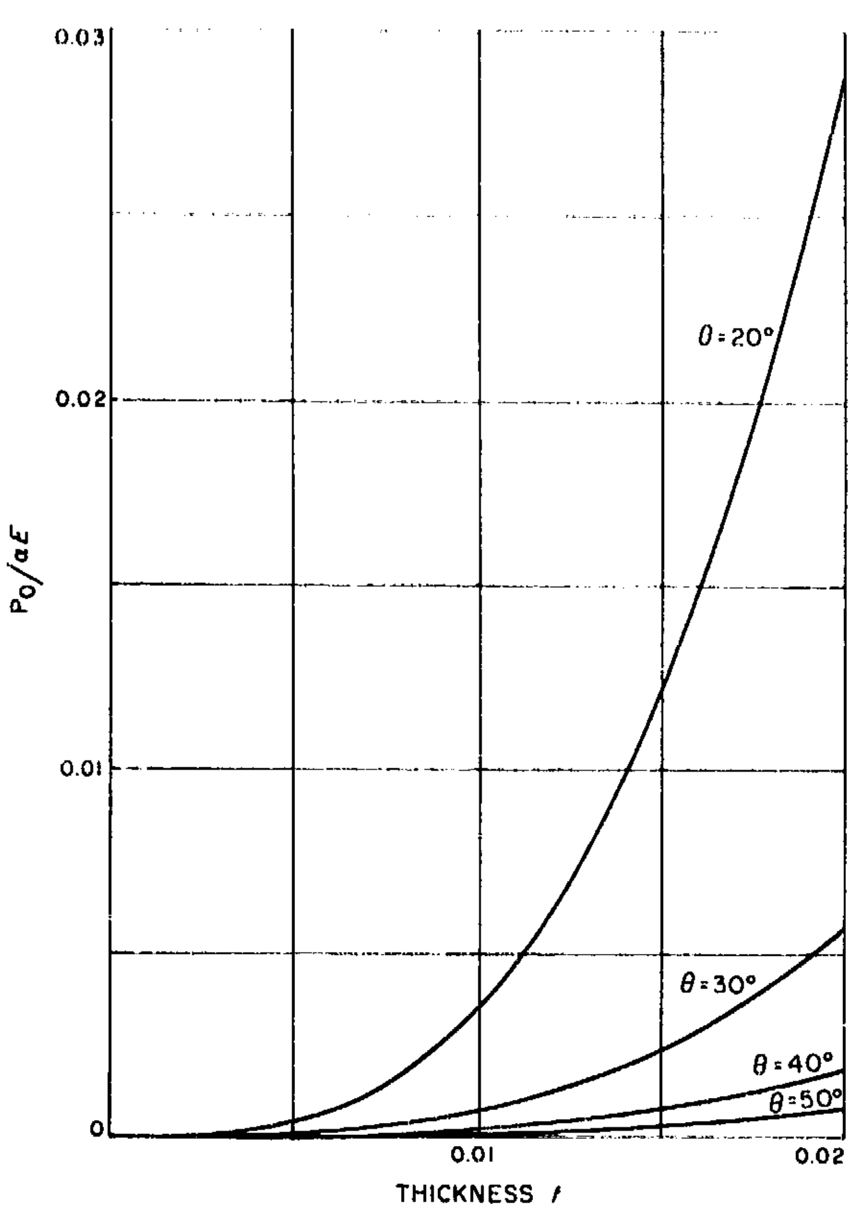


Fig. 12. $P_0/\alpha E$ vs. t for values of θ

provided without being supported by the PG. The space between the metal and PG was set at 0.010 in., the maximum practical braze thickness allowed for a good joint. It was determined that the metal would deform and

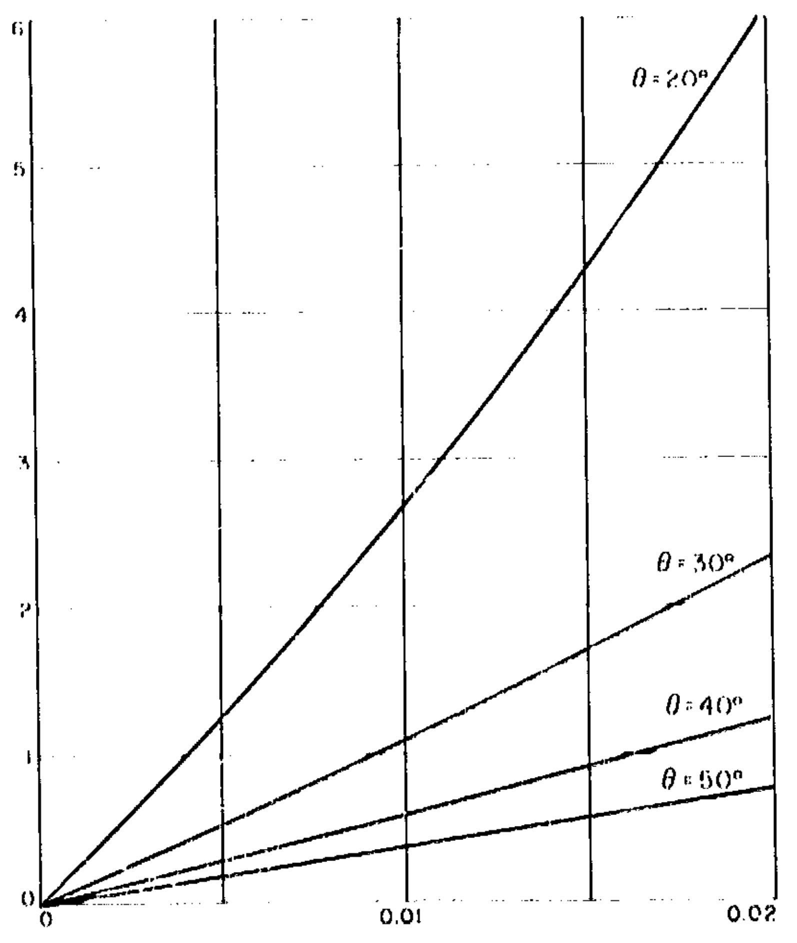


Fig. 13. $\sigma_0/\alpha E$ **vs.** t **for values of** θ

become supported by the PG for all reasonable braze patterns. To overcome this problem, the metal cylinder should have the serpentine pattern stamped into it, creating a raised area necessary to avoid PG support after deformation. The pattern has another advantage: capillary action is greatly decreased in the space-region of the joint, thereby concentrating the braze at the serpentine pattern.

Since yielding occurs, the design procedure for a serpentine-braze joint is an analytical-experimental approach. A test program has been initiated at JPL to evaluate the proposed joint design in which the investigation of tantalum alloys will be emphasized. These alloys exhibit good properties at high temperatures, relative to other refractory metals, and have very good ductility at room temperature. For the 100-lb thrust chamber, the metal cylinder must have sufficient strength to withstand an internal pressure of 150 psi. Also, the shear strength of the braze material is approximately 300 psi. Therefore, as a first joint design for the test program, the metal should be at least 0.008 in. thick with a 1.0-in. PG overlap, combined with a braze pattern spacing of 40° and a braze width of 0.15 in.

C. Structural Analysis System Program Evaluation of Triangular Plate Stiffness Element

T. E. Lang

A structural analysis system (SAS) computer program[1] is being developed that is based upon the direct stiffness method—a recognized accurate analysis technique for complex structures composed of subelements such as bar, beam, torque tube, or membrane elements. In development of the structural analysis system, an attempt was made to develop an element representative of plate and shell structures. The element geometry selected was a triangle in which six degrees-of-freedom are allowed at each of its nodes. In deriving the element stiffness matrix, membrane and bending deformation states were included as well as a stress-strain law allowing representation of isotropic, orthotropic, and aeolotropic materials. This element is being evaluated by use of simple structural configurations to determine its accuracy and convergence properties.

Information to date indicates the element representation is adequate in predicting deflections, slopes, reaction forces, eigenvalues, eigenvectors, and membrane stresses. However, a distribution technique has not been determined that will yield accurately computed bending and transverse shear stresses from the nodal forces. This problem is being studied and constitutes, primarily, the only major inaccuracy associated with the triangular element.

A clamped square plate (Fig. 14) has been studied for which eigenvalues and associated eigenvectors were evaluated. Also evaluated were the deflections, slopes, moments, and shears resulting from a unit normal point load at the center of the plate. For this case, the deflections and slopes are reported.

Nongeometric properties defined for the plate are:

$$E = 10.5 \times 10^6 \quad \text{lb/in.}^2$$

$$\rho = 2.62 \times 10^{-4} \frac{\text{lb/sec}^2}{\text{in.}^4}$$

$$\gamma = 0.333$$

[1]Under development by the Jet Propulsion Laboratory with Philco Corporation, Western Development Laboratories, the principal contractor.

Fig. 15 shows the triangular array selected to idealize the plate. Seventy-two triangles are used, and, with the edges of the plate clamped, twenty-five interior nodes can move. This array can be considered representative of either the complete plate or, because of uniformity, of one-quarter of the plate, assuming symmetrical

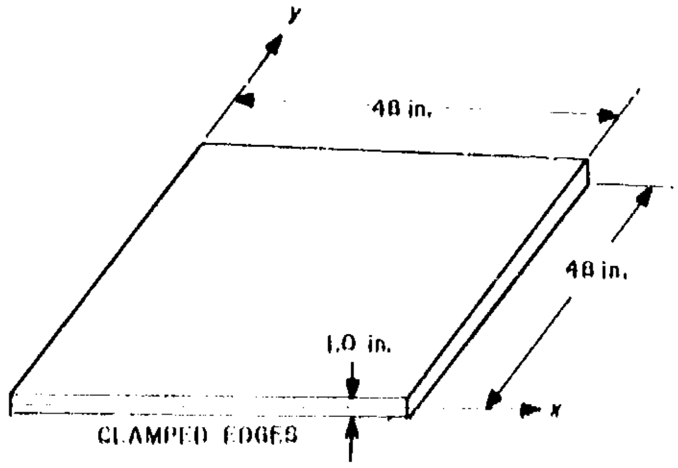


Fig. 14. Plate configuration

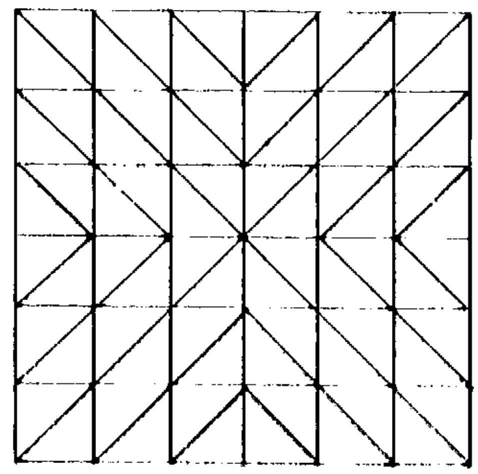
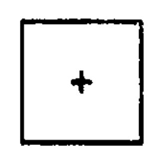
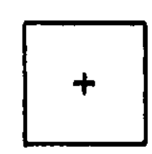
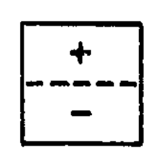
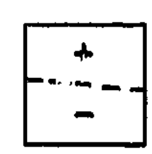
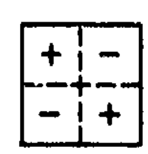
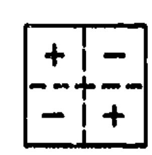
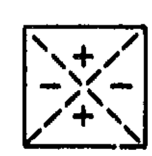
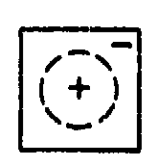
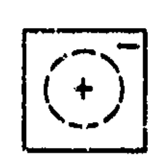
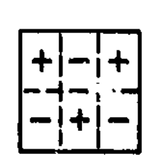
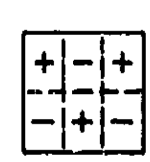


Fig. 15. Triangular idealization

Table 2. Flexural modes and frequencies of a clamped square flat plate

Mode No.	Frequency (exact),[b] cps	Complete plate		Quarter plate[a]		Mode shape	
		Frequency, cps	Error, %	Frequency, cps	Error, %	Exact[b]	Computed
1	153	156	+2.0	156	+2.0		
2	311	350	+12.5	–	–		
3	459	498	+8.5	–	–		
4	558	743	+33.0	571	+2.0		
5	561	705	+26.0	600	+7.0		
6	700	860	+23.0	–	–		

[a] At time of publication, only the symmetric modes had been evaluated.
[b] The exact values are from Ref. 1.

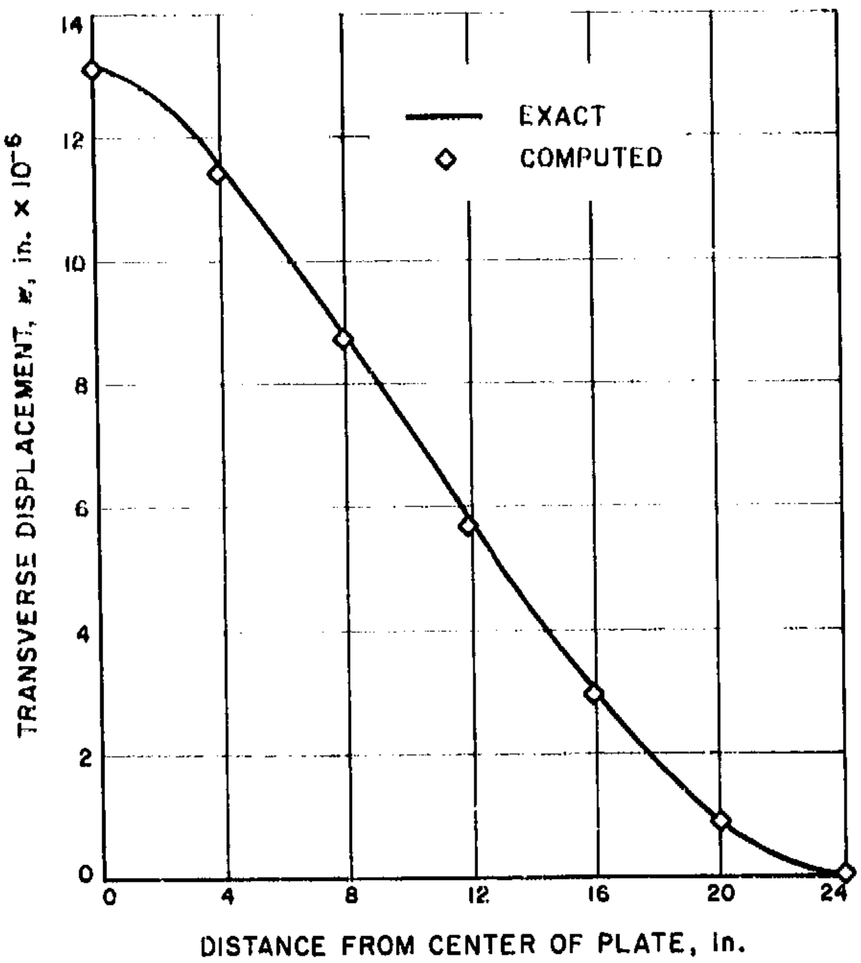


Fig. 16. Transverse displacement of plate from center of plate to center of outer edge—unit normal point load

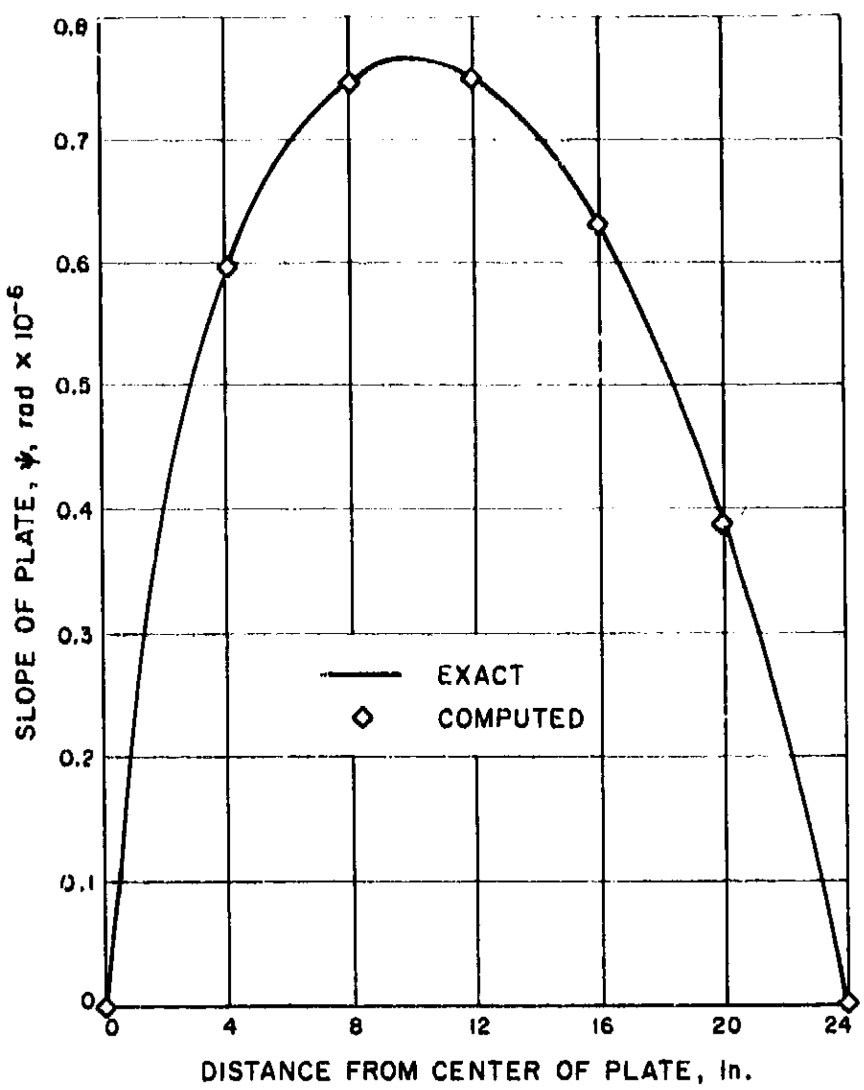


Fig. 17. Slope of plate from center of plate to center of plate edge—unit normal point load

boundary conditions are imposed along two edges. If the array is used to represent one-quarter of the plate, then effectively the representation is that of 109 movable nodes. The mode shapes and frequencies of the low-frequency flexural modes of the plate were computed (Table 2) using both the complete- and quarter-plate representations.

The mass distribution technique materially changes the accuracy of computed frequencies. The values reported in Table 2 were computed using a mass distribution corresponding to a complimentary energy formulation of the problem. This distribution, coupled with the potential energy formulation of the stiffness matrix, can result in computed frequencies less than exact values. If a potential energy formulation is used throughout, then all computed frequencies should be greater than corresponding exact values. Except for the second mode, computed mode shapes were identical to exact shapes. The computed second mode nodal line curved slightly—whereas in the exact node, the nodal line parallels two sides of the plate and divides it into two equal parts. The cause of this variation is not known.

The deflections of the plate, due to a central unit normal point load, were computed using the quarter-plate representation. The transverse displacement w and slope $\psi = \partial w/\partial x$ along an axis perpendicular to a side of the plate and passing through the center of the plate are plotted in Figs. 16 and 17. For the rough grid, the deflections and slopes closely approximate the exact values.

Reference

1. Harris, C. M., and Crede, C. E., "Shock and Vibration Handbook," Vol. 1, Chapter 7, McGraw-Hill Book Co., Inc., New York, 1961.

ENGINEERING FACILITIES DIVISION

X. Instrumentation

A. An Infrared Photometer for Measuring Freon 13 Density in the Exhaust of a Hydrogen–Freon 13 Gas Core Reactor Model

G. A. Wiker and R. C. Willson

1. Introduction

The testing of gas core reactor models includes the use of Freon 13 to simulate a heavy reacting gas. An indication of the effectiveness of a given model configuration is the amount of Freon 13 appearing in its exhaust. An infrared photometer has been constructed to measure quantitatively the amount of Freon 13 in the exhaust by detecting its absorption of infrared radiation.

Freon 13 has a strong absorption line centered at the wavelength: $\lambda = 8.26\mu$. The profile of this absorption feature has a width and depth that increase with rising Freon density. The infrared radiation reaching the photometer detector is restricted to a narrow wavelength range just off the absorption line center on the steep slopes of the family of absorption profiles corresponding to the Freon 13 densities of interest. A maximum rate of change in intensity is then seen by the detector as the density of Freon 13 in the radiation path varies.

2. Photometer

A schematic drawing of the radiometer is seen in Fig. 1. The infrared source is a cone-shaped, ceramic heater element of 1000-w capacity operated at a regulated 110 v ac. This mode of excitation is satisfactory since the ceramic interior of the source, to which the detector field of view is limited, is massive. A mirror, 7.6 cm in diameter and 45.7 cm focal length, gathers the source energy and directs it along the photometer axis. The infrared beam passes through the water-cooled aperture tube, which restricts the field of view for the downstream optics to the concave mirror. The emergent radiation, after modulation at a frequency of 230 cps by a rotary chopper, encounters the narrow band-pass filter.

The filter is a two-element device comprised of multilayer interference filters deposited on two germanium discs. The first is a blocking filter that reduces transmission to $\leq 0.5\%$ for wavelengths: $\lambda \leq 8.01\,\mu$ and $8.65\,\mu \leq \lambda \leq 14.00\,\mu$. The Irtran II windows used elsewhere in the

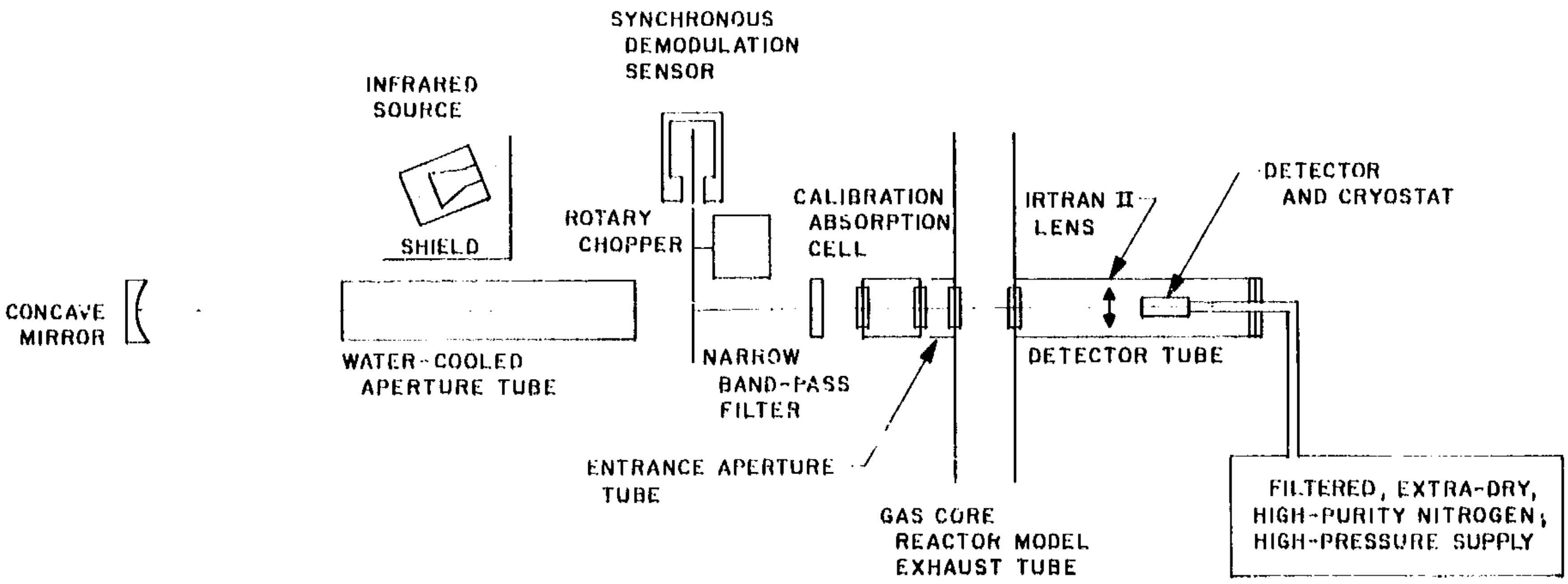


Fig. 1. Schematic diagram of the 8.26-μ infrared photometer

instrument do not transmit appreciably beyond 14 μ and thereby complement the blocking function of the filter. The second germanium substrate has a multilayer filter deposited on it with a transmission of 51% at 8.26 μ decreasing to 25% for a 0.18-μ bandwidth.

An absorption calibration cell can be placed in the optical path between the filter and the exhaust tube. This cell, into which Freon 13 can be introduced at any desired pressure, duplicates the optical path length of the exhaust tube. Low reflection coated Irtran II windows are employed as the optical ports in the cell. The optical ports of the exhaust chamber are also low reflection coated Irtran II windows. The unabsorbed radiation emerging from the exhaust chamber is focused by a 2.5-cm focal length Irtran II lens to form a 2.2-mm diameter image of the source on the detector element.

The detector is a Philco GPC 201A gold-doped germanium type with a 2.0-mm square germanium flake. The detector is operated at 77°K using a Philco C-100 Joule-Thompson liquefier that mates with the built-in Dewar of the detector. Extra-dry, high-purity nitrogen is passed through a Walter Kidde and Co. Type 840873 chemical dryer and filter, and supplied to the C-100 liquefier at 900 psi for continuous operation. Higher starting pressures are used for rapid cool-down from ambient temperatures.

During operation of the reactor model, expanding hydrogen in the exhaust chamber causes sufficient cooling to condense atmospheric water vapor onto its exterior surfaces. To prevent condensation on the outside surfaces of the Irtran II optical ports, the detector chamber and entrance aperture tube are continuously flushed with dry nitrogen. Fig. 2 shows the optical system.

3. Transport of Radiant Energy

The flux of radiant energy incident upon the detector can be expressed as:

$$\Phi = TS\Omega B \tag{1}$$

where

Φ = flux, w

T = transmission factor (unitless)

S = limiting source area (cm^2)

Ω = limiting admissible solid angle (sterad)

B = radiant intensity of source (w-cm^{-2}-sterad^{-1})

T is the net transmission factor of the instrument. It accounts for the imperfect transmission and reflection of various elements of the system. Both calibration and normal operation of the system involve the same number and types of optics. T for the photometer is 0.2.

The quantities S and Ω can be combined to form as their product the etendue (extent) of the system.

$$U = S\Omega \qquad \text{Etendue} \tag{2}$$

For this instrument the etendue is: $U = 1.3 \times 10^{-2}$ cm^2-sterad.

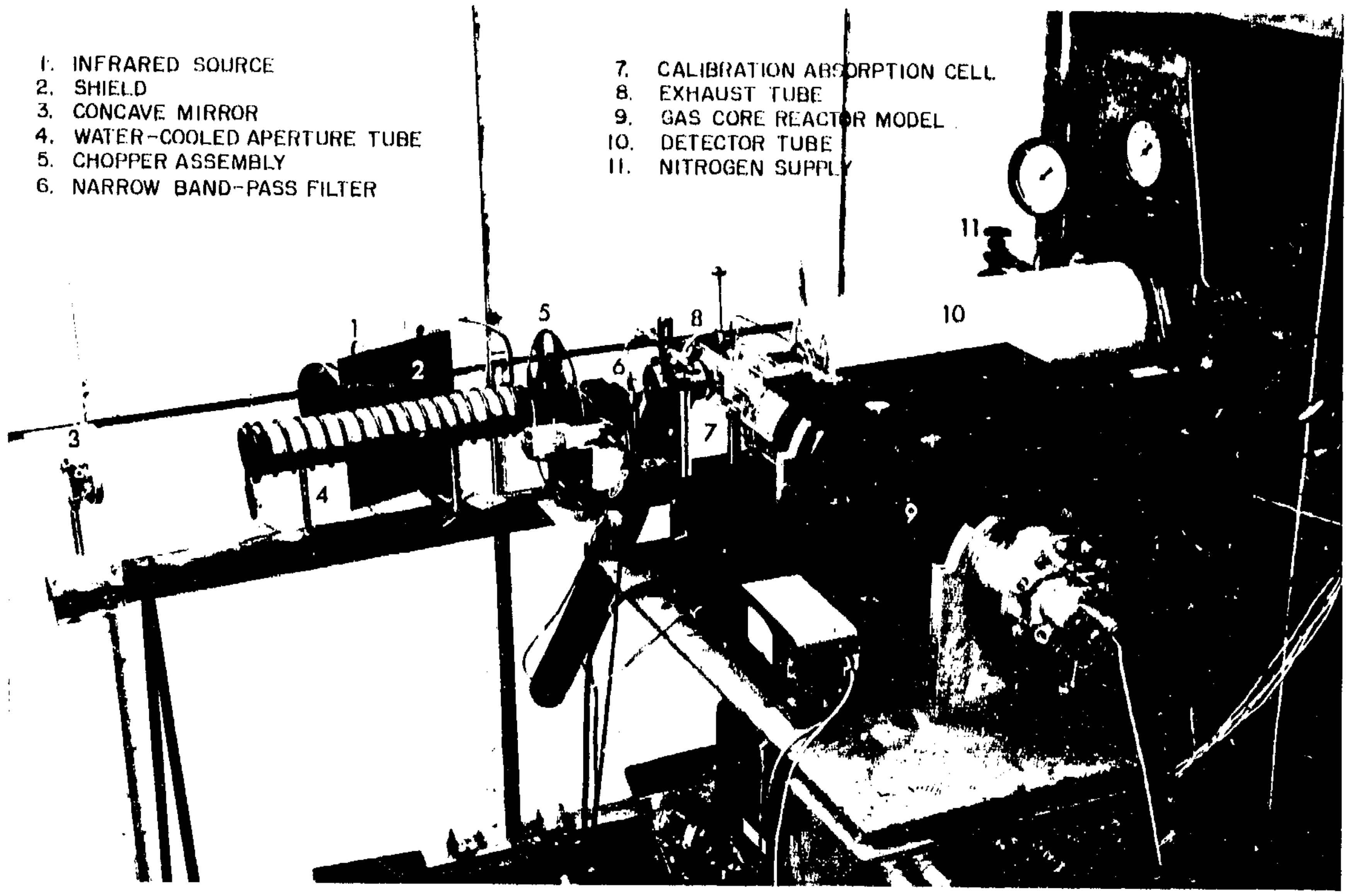


Fig. 2. Infrared photometer

B, the radiant intensity of the source, is that corresponding to a 1500°K black body. $B = 28.75$ w-cm^{-2}-sterad^{-1}. We will define an effective source, B_{eff}, by applying the narrow band-pass filter function to B:

$$B_{eff} = \int_{8.17\mu}^{8.35\mu} W(\lambda)\, d\lambda. \tag{3}$$

This yields a $B_{eff} \cong 10^{-3}$ w-cm^{-2}-sterad^{-1}.

The radiant flux incident upon the detector is then, by Eq. (1):

$$\Phi \simeq 3 \times 10^{-6} \text{ w}. \tag{4}$$

4. Electronics

The infrared detector signal amplification system (Fig. 3) consists of a high-gain, low-noise, band-pass amplifier which amplifies the modulation components of the infrared detector output signal. This is followed by a synchronous demodulator, amplifier, and low-pass filter which produce the electrical signal proportional to the infrared radiation incident on the detector.

The band-pass amplifier is made up of three units; the first is located as close as possible to the infrared detector. It provides 60 μa of DC excitation to the variable resistance infrared detector and impedance transformation and amplification for the detector output signal. The amplifier consists of a field-effect input transistor (2N3578) chosen for its low noise characteristic, and a boot strap-connected silicon output transistor. The input impedance is 22 MΩ and the equivalent input noise is 4.5 μv RMS in a 1-kc bandwidth. The output impedance is 400 Ω and the signal gain is +2.

The second unit is a Tektronix Type 122 differential input amplifier. It provides variable high-pass and low-pass filters and switch-selected voltage gain of 100 and

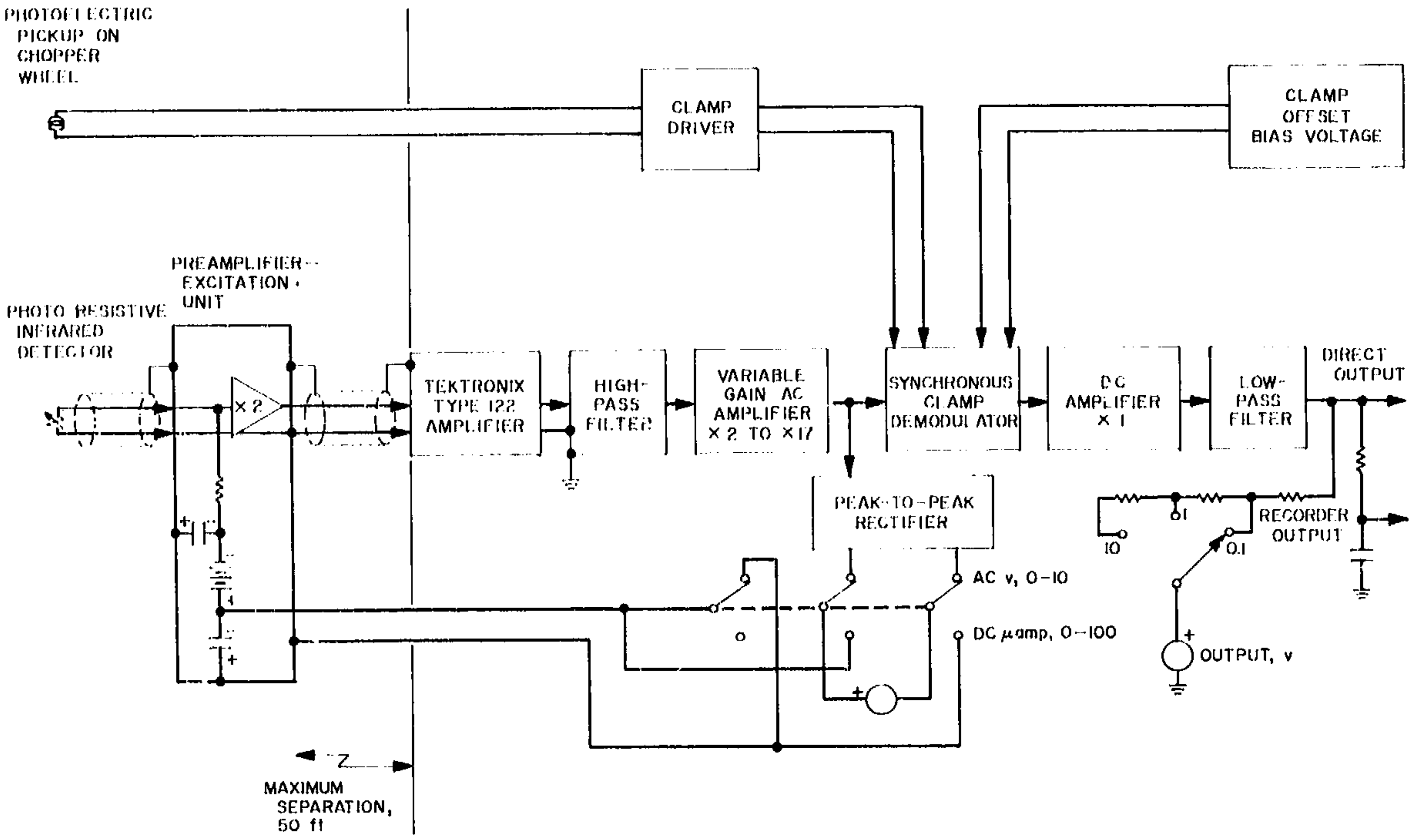


Fig. 3. Infrared signal amplification system

1000. At a gain of 1000, the Type 122 amplifier has an equivalent input noise of 0.8 μv RMS in a 1-kc bandwidth.

The third unit contains a high-pass filter with a sharp low-frequency cutoff starting at 100 cps to lower the system noise, followed by an AC negative feedback amplifier with a voltage gain control adjustable from 2 to 17. A 0- to 10-v range peak-to-peak AC voltmeter monitors the output of the variable gain amplifier, and can be switched to monitor the excitation current of the infrared detector on a 0- to 100-μa range. The variable gain AC amplifier output is fed to the synchronous clamp demodulator which restores the DC component of the infrared proportional signal. The synchronous clamp is driven by the amplified and shaped output of a photoelectric detector mounted on the infrared source chopper wheel assembly. An adjustable negative offset bias voltage allows expanded scale operation by bucking out up to 8 v of the clamp demodulator output. The partially filtered output of the clamp demodulator is connected to a gain of one, impedance transforming, DC amplifier. This drives a low-pass filter with a sharp high-frequency cutoff starting at 75 cps which removes the modulation components of the infrared proportional signal. The output signal is monitored with a DC meter. The three switch-selected meter ranges are 0 to 10, 0 to 1, and 0 to 0.1 v peak. The direct output signal is nominally set to between +2 and +4 v. An attenuated, filtered output with a nominal signal level of +20 to +40 mv and a band-pass of 1 cps is provided for strip chart recorders.

XI. Aerodynamic Facilities

A. Wind Tunnels

J. T. Welton, R. G. Harrison, Jr., J. E. Marte, R. W. Weaver, P. Jaffe, R. H. Prislin, R. D. Wood, and R. J. Hemstead

1. Free-Flight Telemetry, *J. T. Welton and R. G. Harrison, Jr.*

To develop a free-flight telemetry capability for the JPL wind tunnels, Test 21-182B was conducted in the 21-in. Hypersonic Wind Tunnel. Support-free base pressure data are being obtained from 10- and 15-deg half-angle cones at Mach numbers 6.0 and 8.0. Pressures at the base of the free-flight model are sensed by a small telemetry package (Fig. 1), and transmitted to a receiving and recording system outside the wind tunnel test section. A representative data point copied from an oscilloscope trace is shown in Fig. 2.

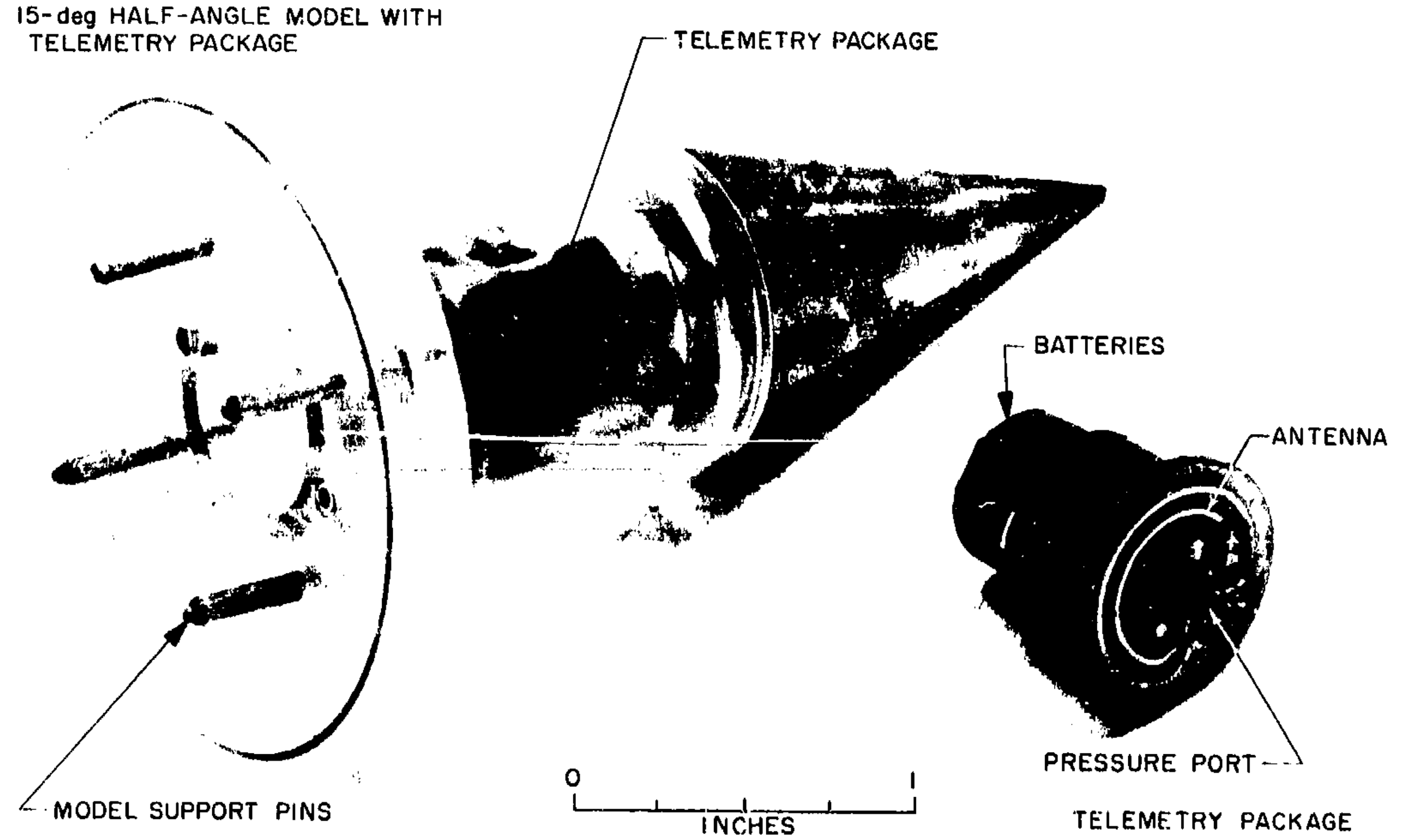


Fig. 1. Telemetry package and model

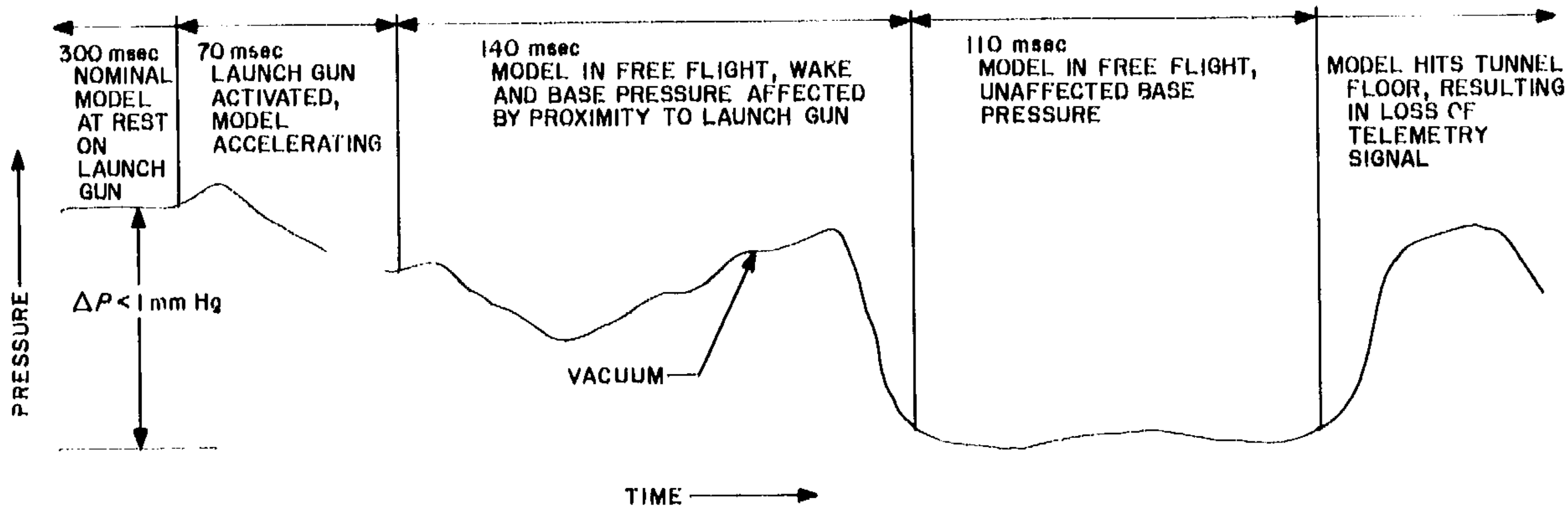


Fig. 2. Sample oscilloscope trace free-flight telemetry base pressure

Using 10 different telemetry packages, 120 flights have been made. Some packages have been flown as many as 33 times with only minor repair, such as replacing batteries following flight. Approximately 25% of the flights have produced meaningful data, the remaining 75% producing useful system development information. Models are launched into free flight in the wind tunnel test section by a pneumatic launch gun (*SPS 37-31*, Vol. IV, p. 135). Data are recorded while the model is in the test section, and the model is recovered by a wire net at the tunnel floor. Models usually can be re-flown at once. All flights are photographed with a 16-mm Fastax movie camera to determine model attitude during flight.

To further the capability in free-flight telemetry, future plans include telemetry of model base and surface temperatures.

2. Langley 20-ft Free-Spinning Wind Tunnel Dynamic Stability Tests of Mars Entry Configurations, *J. E. Marte and R. W. Weaver*

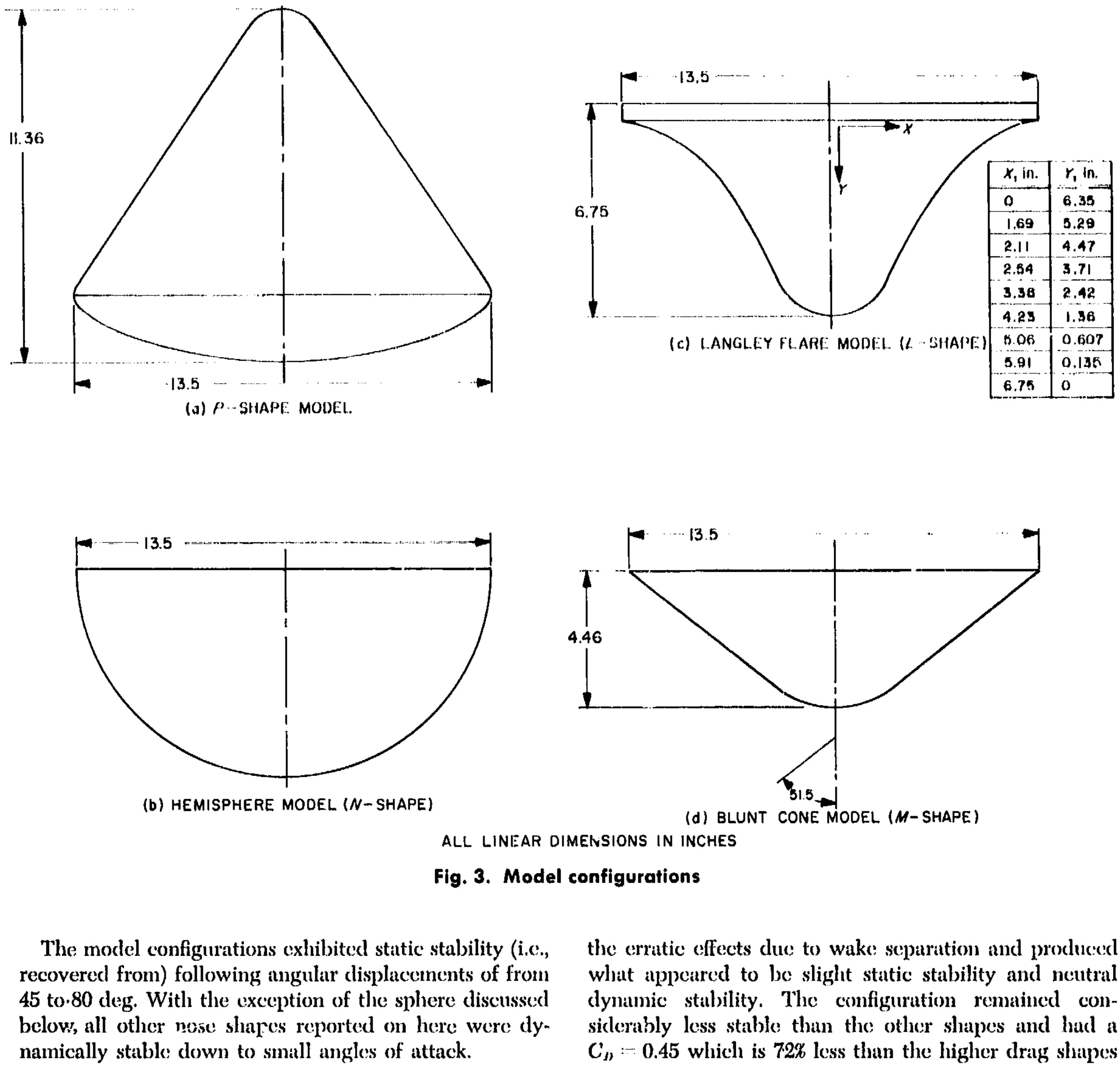
Tests were run in the 20-ft Free-Spinning Tunnel at Langley Research Center to investigate the low subsonic ($V < 90$ ft/sec) dynamic stability of a series of possible Mode 1 entry configurations. This tunnel has a vertical test section with air at atmospheric supply pressure flowing upward, allowing models to be flown in steady free flight by adjusting the air velocity so that the model's drag equals its weight. Data are recorded on motion picture film which shows, in addition to the model's attitude as a function of time, the tunnel flow velocity.

Details of the geometry of the four nose configurations tested are given in Fig. 3; the center-of-gravity locations and pitch moments of inertia are given in Table 1, along with the static coefficients C_D and C_{m_α} which could be computed from the experimental data. All models were axially symmetric. Plane surfaces, hemispheres, and cones were used as afterbody shapes. The hemisphere used had a diameter which was 70% of the maximum model diameter and the cone was geometrically identical to the cone nose shape. Moments of inertia were made as large as possible, consistent with the model size, construction, center-of-gravity location, and limitation on total weight. This results in a rotational similarity parameter ($\rho d^5/I$) for these data of about 100 times the Mars value; center-of-gravity locations were, in general, as far forward as physically possible.

Other afterbodies were tested with the four nose shapes, in addition to those shown in Table 1, but no data suitable for numerical analysis were obtained. As shown for the three flare nose configurations, the drag coefficient and pitch moment slope are principally functions of the nose shape. The values given may be considered as representative for any configuration having that nose shape and a definite separation point as far as these parameters are concerned.

Table 1. Data summary

	L	L + cone	L + hemisphere	M	N	Units
cg/D	0.245	0.269	0.269	0.195	0.210	
I_y	13.60	10.96	13.13	9.52	8.99	Slug–ft² × 10⁵
C_D	1.09	1.10	1.11	0.88	0.45	
C_{m_α}	0.11	−0.15	—	−0.18	0.36	Per rad



X, in.	Y, in.
0	6.35
1.69	5.29
2.11	4.47
2.54	3.71
3.38	2.42
4.23	1.36
5.06	0.607
5.91	0.135
6.75	0

ALL LINEAR DIMENSIONS IN INCHES

Fig. 3. Model configurations

The model configurations exhibited static stability (i.e., recovered from) following angular displacements of from 45 to 80 deg. With the exception of the sphere discussed below, all other nose shapes reported on here were dynamically stable down to small angles of attack.

A smooth 13.5-in. diameter sphere with $cg/D = 34.7\%$ moved in an erratic path apparently due to wake separation effects. It appeared to be neutrally stable statically and slightly unstable dynamically. The drag coefficient was 0.40. Large-scale roughness was introduced on the sphere surface in the form of six rings of 3/32-in. rod, 7½ in. in diameter, attached to the surface so that the center of each ring was located at 90 deg from each of its immediate neighbors. This roughness considerably reduced the erratic effects due to wake separation and produced what appeared to be slight static stability and neutral dynamic stability. The configuration remained considerably less stable than the other shapes and had a $C_D = 0.45$ which is 72% less than the higher drag shapes tested. Other related shapes tested included a "stepped" sphere having an afterbody hemisphere 70% of the forebody hemisphere. The stability characteristics appeared to be about the same as the hemisphere-flat configuration of Table 1 but no data suitable for numerical analysis could be obtained.

Data for the P-shape (*Apollo*-like shape) are classified and will appear in a JPL technical report (to be published). The report will include numerical values for the dynamic damping parameter, $\overline{C_{m_q}}$.

Early in the test program two stabilizers of types which had been effective in earlier drop tests with low moment of inertia models (stiff wire and thread) were tried on the P-shape. They appeared to be completely ineffective in stabilizing a body of the size and inertia of those under test, so stabilizers of this type were eliminated from the test program pending further investigation of the scaling and interaction problems involved. Some small parachutes were tested as stabilizers and for the most promising shapes they were able to stabilize the model from an initial attitude of 180 deg with chutes as small as 3 in. in diameter.

On the basis of observations made during this test program, several general impressions were gained. None of the high-drag, low-fineness-ratio shapes now being considered was able to recover from a tumbling motion under the test conditions. Those shapes which had a positive point of flow separation, such as a step or large angle of expansion near the point of maximum diameter and an afterbody shape which did not permit reattachment, exhibited the most satisfactory stability characteristics.

3. Dynamic Stability Research, P. Jaffe and R. H. Prislin

For the past two years, a series of dynamic stability tests has been conducted in both the supersonic and hypersonic wind tunnels. The purpose of the program has been four-fold: (1) to develop and improve dynamic stability testing techniques; (2) to investigate thoroughly the dynamic stability characteristics of 10-deg cones; (3) to determine the effect of blunting 10-deg cones; and, (4) in general, to obtain a better understanding of the dynamic stability decay mechanism.

This progress report presents results from more recent tests and also, summary information from earlier tests. Only a portion of the data from the recent tests is presented and this is considered preliminary. The principal test parameters were Mach number, Reynolds number, oscillatory frequency, oscillatory amplitude, center-of-rotation, and bluntness ratio. Two distinct testing techniques were utilized in this program: the free-flight technique and the sting-supported free-oscillation technique. The free-flight technique and data reduction are described in *SPS 37-24* and *37-31*, Vol IV, and Refs. 1 and 2. The free-oscillation technique and data reduction are described in Ref. 3. Table 2 contains a summary of the tests and the corresponding test parameters. The nomenclature of Table 2 is explained in Fig. 4 which shows a typical model, with both a flat and a contoured base and a bluntness ratio, r (nose)/d (base), of 0.094. For all blunt cones the center-of-gravity location is referenced in lengths of a hypothetical sharp cone from the hypothetical sharp nose. The free-flight models ranged from 0.5- to 1.0-in. diameter; the free-oscillation models were about 5.25 in. in diameter.

Figs. 5 and 6 show the effects of oscillation amplitude on the dynamic stability coefficient. Fig. 5 contains free-oscillation and free-flight data for the sharp cone configuration at several centers of rotation. The free-flight data, shown as dashed lines, were reported on in *SPS*

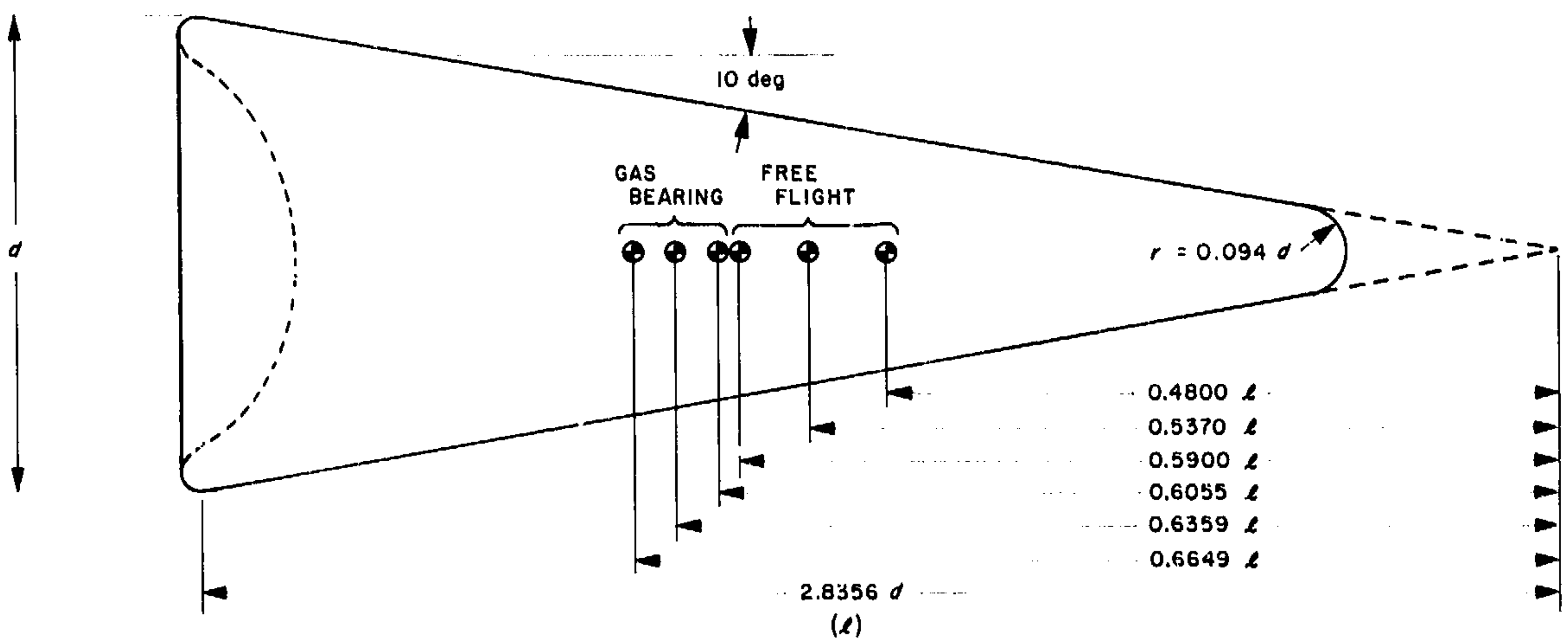


Fig. 4. Typical test configuration

37-24, Vol. IV and Ref. 1. It should be noted that the free-flight data were obtained at $M_\infty = 4.5$ and the free-oscillation data at $M_\infty = 6.0$. Both techniques indicate the same trend of increasing dynamic stability with increasing amplitude. The free-oscillation data begin to rise somewhat earlier (4 to 6 deg) than do the free-flight data (8 to 10 deg) but, thereafter the curves appear to remain parallel. The effect of amplitude on a blunted 10-deg cone is shown in Fig. 6. These data, taken at $M_\infty = 6.0$ by the free-oscillation technique, indicate a greater change in the dynamic stability coefficient with amplitude than was seen with the sharp cone. For the 0.6 l case, there is a 14% rise in the dynamic stability as the amplitude increases from 4 to 18 deg with the sharp cone whereas, there is a 57% increase with the blunted cone. It should be noted, however, that in case of very blunt cones ($r/d = 0.43$) this trend reversed itself to the point where there is no discernable amplitude effect (Ref. 3).

Fig. 7 presents the effect of center-of-rotation on the dynamic stability coefficient at constant amplitude for the

Table 2. Test parameters

Configuration	Test No.	Mach No.	cg (% l aft of sharp nose)	R_e/in. × 10⁻⁶	wd/v × 10²
A. Free oscillation					
Sharp cone	21-176	6.0–8.0	60.40–66.18	0.04–0.20	0.25–0.55
$r/d = 0.091$	21-176	6.0–8.0	60.40–66.18	0.07–0.20	0.36–0.68
$r/d = 0.094$ contoured base	21-176	6.0–8.0	60.55–66.49	0.02–0.28	0.17–0.53
B. Free flight					
Sharp cone	21-172 20-555 20-562 20-585	2.0–4.5	49.4–61.4	0.17–0.34	0.94–1.99
$r/d = 0.094$ contoured base	21-172	4.0–8.0	48.0–59.0	0.22–0.25	0.57–1.19
$r/d = 0.1$	20-598	4.0	54.7	0.28	1.44
$r/d = 0.2$	20-598	4.0	66.0	0.28	1.13
$r/d = 0.3$	20-598	4.0	66.0	0.28	1.19
$r/d = 0.4$	20-598	4.0	81.7	0.28	1.18

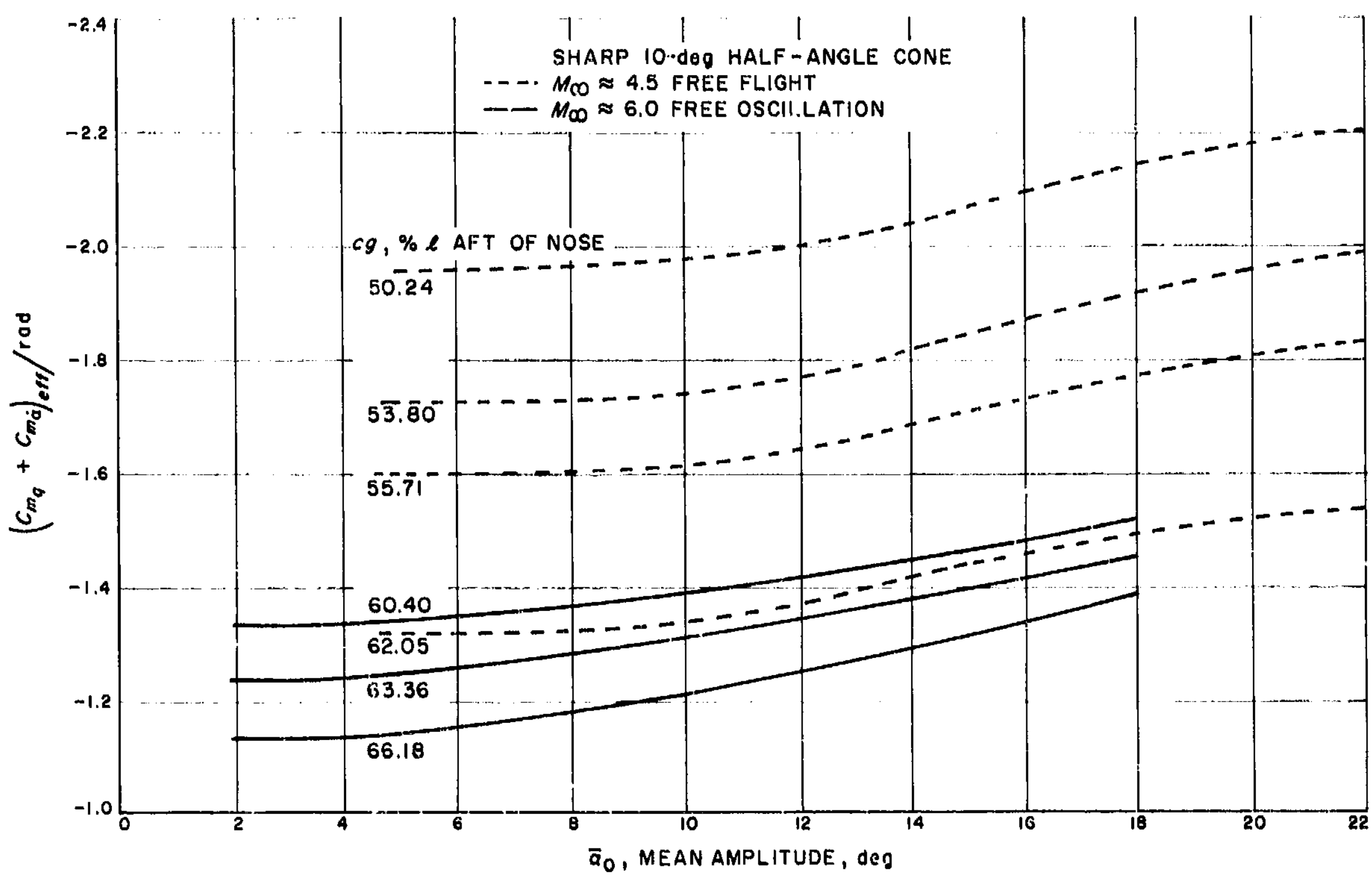


Fig. 5. Effects of oscillation amplitude on dynamic stability for a sharp cone

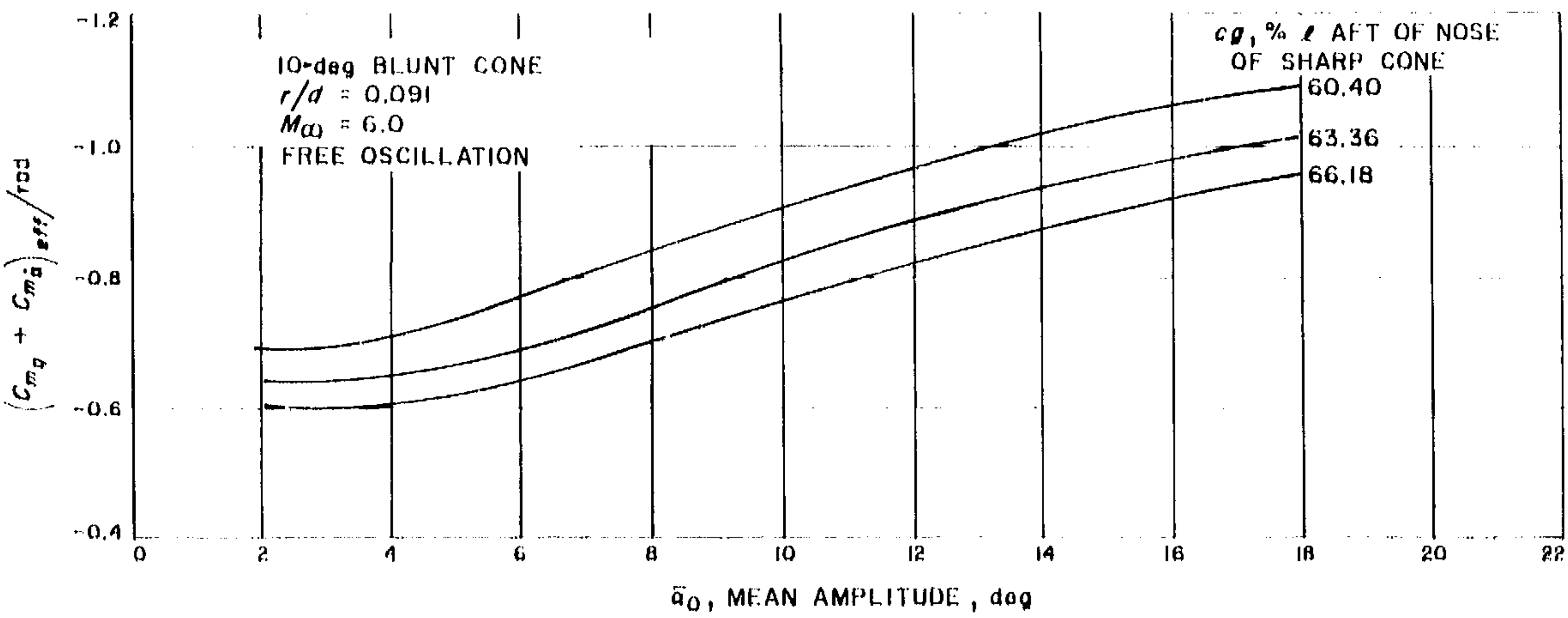


Fig. 6. Effects of oscillation amplitude on dynamic stability for a blunted cone

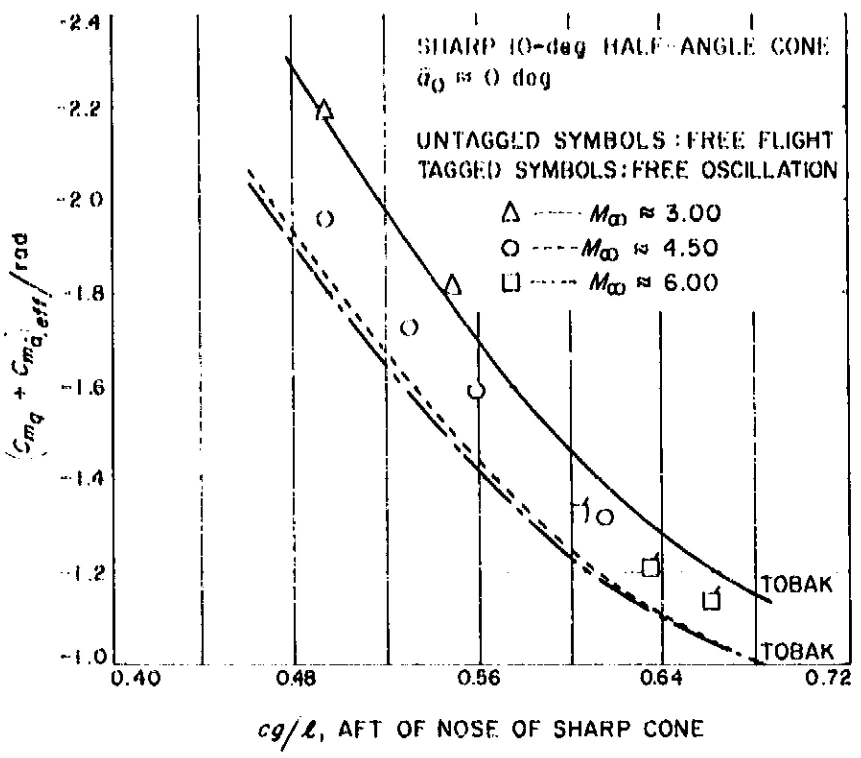


Fig. 7. Effects of center-of-rotation on dynamic stability for a sharp cone

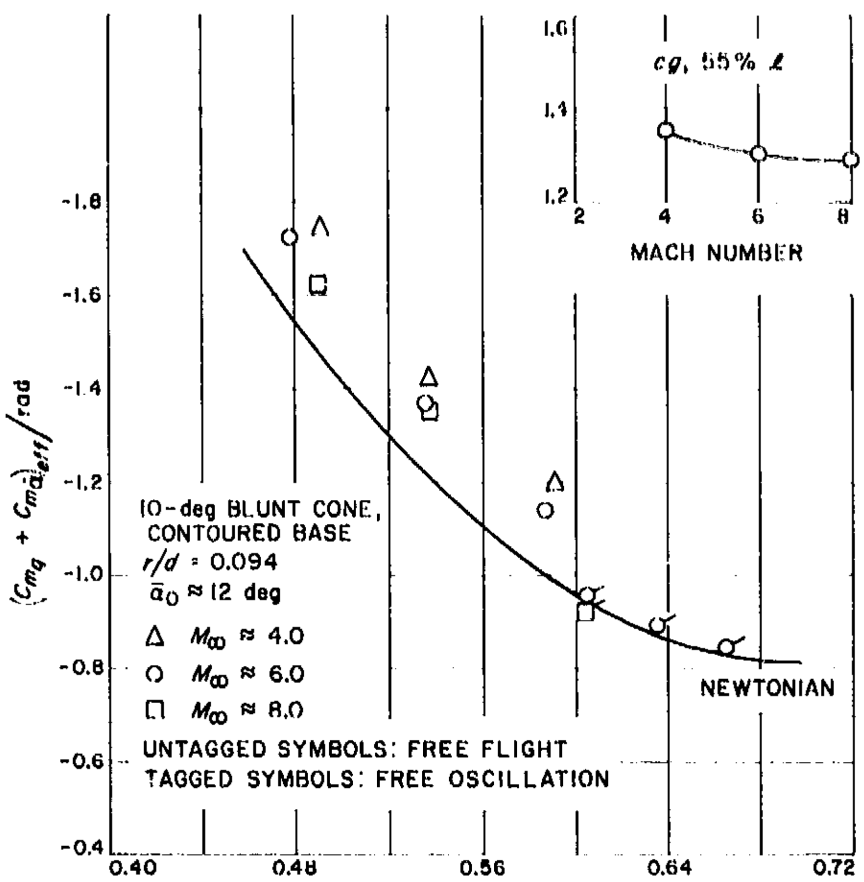


Fig. 8. Effects of center-of-rotation on dynamic stability for a blunted cone

sharp cone. This figure includes free-flight data taken at $M_\infty = 3.0$ and 4.5 (Ref. 1) and free-oscillation data taken at $M_\infty = 6.0$. The data are compared with Tobak's first- and second-order potential flow solution, applicable near 0 angle of attack (Ref. 4). There is good agreement at all Mach numbers, the maximum deviation being on the order of 7%. This plot shows excellent correlation between the data obtained by the two alternate testing techniques.

The same type of data for a blunt 10-deg cone, $r/d = 0.094$, with a contoured base are shown in Fig. 8. Again there are several Mach numbers and both free-flight and free-oscillation data are presented. The data are compared with a curve calculated from the Newtonian impact theory. The free-flight and free-oscillating curves exhibit the same shape; however, the levels of the two curves differ by about 10%. No explanation for this inconsistency is apparent at this time. In the upper right

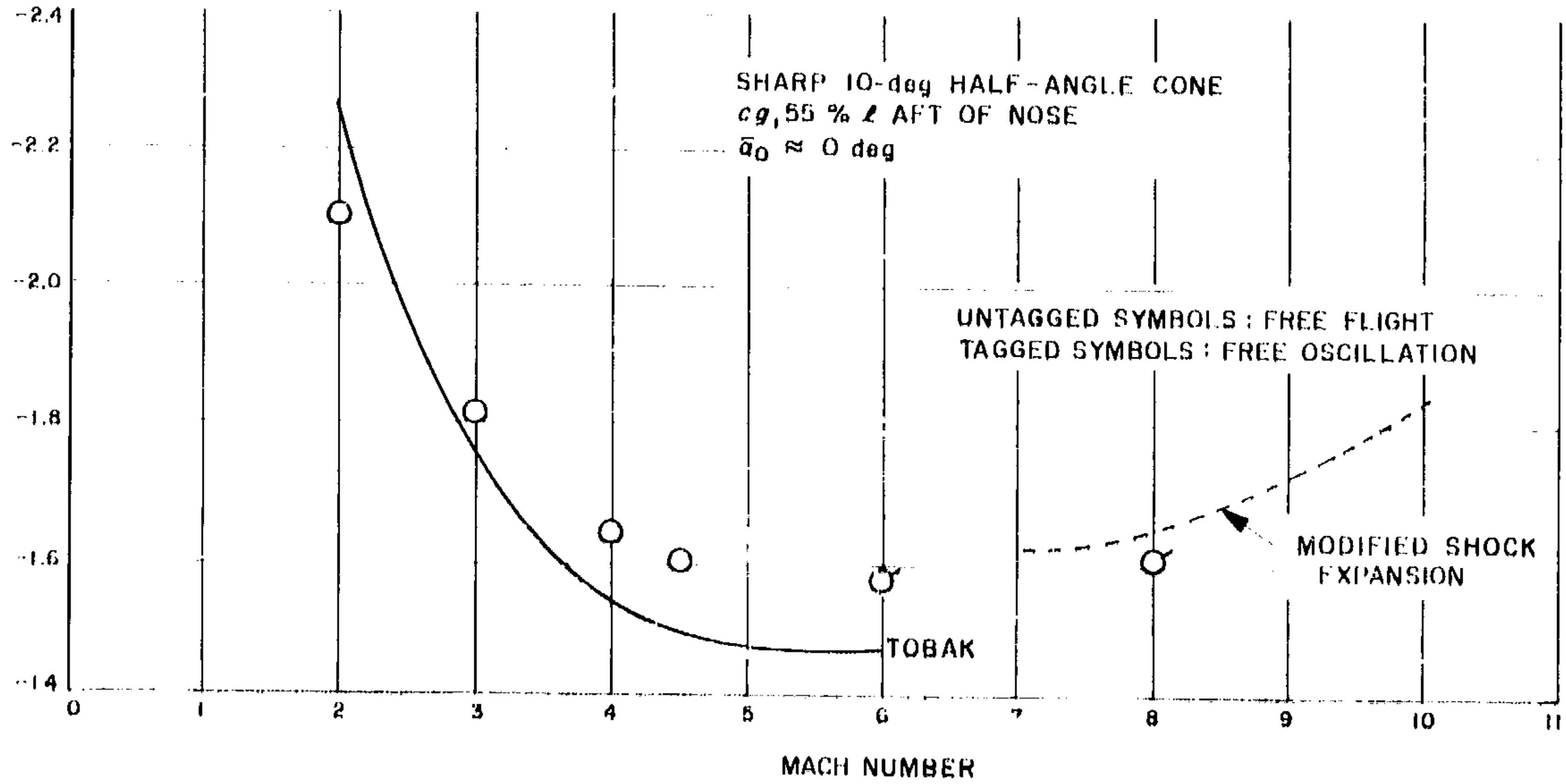


Fig. 9. Effects of Mach number on dynamic stability for a sharp cone

of Fig. 8 is a summary of Mach number effect for this blunt configuration, obtained from free-flight data.

Fig. 9 shows the effect of Mach number on dynamic stability for the sharp cone. Again, both free-flight and free-oscillation data are shown on this plot. The solid line represents Tobak's potential flow solution while the dashed line represents a shock expansion solution (Ref. 5). Both theories agree with experimental data to better than 7%. Both theory and experimental show a pronounced variation in the coefficient in the region of $M_\infty = 2$ to 3, and both show an indication of an increase in dynamic stability in the $M_\infty = 6$ to 8 region.

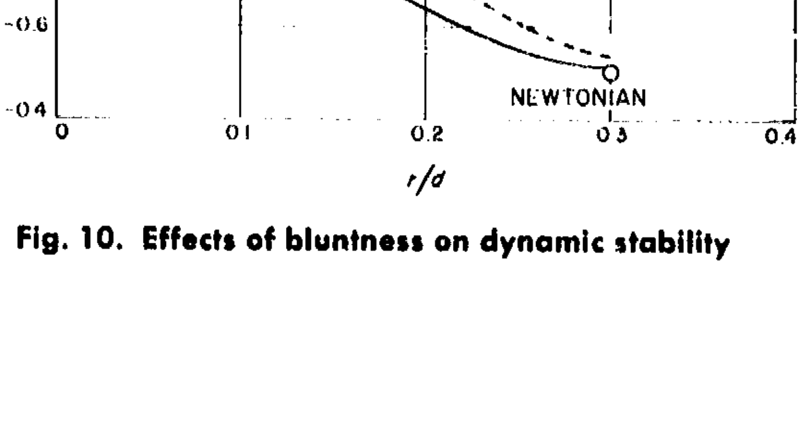


Fig. 10. Effects of bluntness on dynamic stability

The effects of bluntness on dynamic stability from the free-flight data are shown in Fig. 10 along with a curve from the Newtonian impact theory. Points at $r/d = 0.2$ and 0.3 are preliminary data not as reliable as the remaining points. Experimental data for the sharp cone ($r/d = 0$) are about 30% above Newtonian. As the bluntness ratio increases, experimental and Newtonian curves tend to converge, meeting near $r/d = 0.3$. Experimental data at $r/d = 0.43$ (Ref. 3) also indicate good agreement with Newtonian, confirming the general trend.

4. *Free-Flight Capsule Afterbody Studies*, R. W. Weaver

The stability characteristics of afterbodies during entries at angles of attack in the 180-deg region are important factors in the choosing of an over-all entry shape. Wind tunnel studies were made to investigate the effect that the internal configuration of cylindrical afterbodies has on the stability during rearward entry. It is mandatory for an entry capsule to orientate itself in a heat shield forward position prior to maximum heat pulse to prevent burnup.

These tests were conducted using cylindrical afterbodies of three different lengths with six different internal configurations. The models were constructed of magnesium with copper inserts in the forebodies to insure that the centers of gravity were as far forward as possible. The models are shown in Fig. 11. Each forebody–afterbody combination is shown with the various inserts that

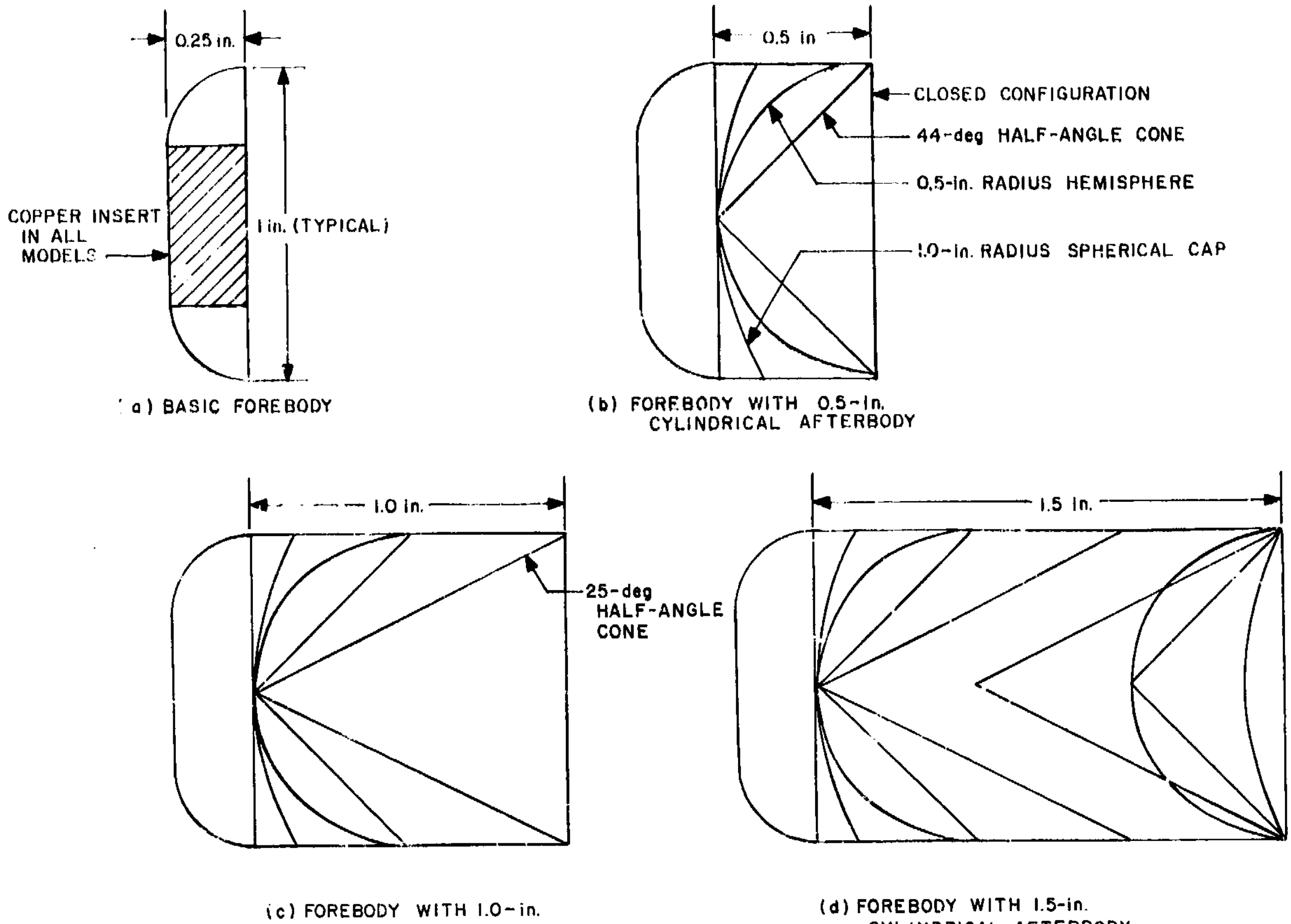


Fig. 11. Forebody–afterbody combinations tested

were used with that particular combination. The models were wire-launched in the 20-in. Supersonic Wind Tunnel at angles of attack of 180 and 150 deg. This method of testing was used in order to obtain interference-free data for the models. The Mach number was 4.00 and the dynamic pressures were 123 and 154 lb/ft^2. These tunnel conditions were picked because they were the lowest available which would yield readable high-speed motion pictures with schlieren effects. The motion pictures and the tunnel conditions were the only data acquired. The only data reduction that was anticipated and accomplished was a visual evaluation of the movie data.

The forebody alone was stable at 180 deg. When launched at 150 deg, the forebody oscillated several times but was damping to 180-deg position.

When the forebody with the 0.5-in. cylindrical afterbody was launched at 180 deg, it would stabilize at 180 deg. The insert configuration made little or no difference. Because this model could not be launched successfully at 150 deg, no evaluation of it can be made.

The forebody with the 1-in. cylindrical afterbody at 180 deg flipped right after launch, which was due to the supporting wire catching on the model after release. This was not known until the films were viewed and thus no check runs were made. All of the runs of this configuration at 180 deg had problems with the wire releasing properly. This model flipped to, and would become stable at, zero angle of attack when launched at 150 deg.

The forebody with the 1.5-in. cylindrical afterbody appeared to flip at both angles of launch. This model also experienced the wire-release problems.

The data indicate that the internal configuration of the cylindrical afterbodies has a negligible effect on the performance of the afterbody. The main effect is the length of the afterbody.

During evaluation of the data films several peculiarities in the shock patterns were noted. In one instance the shock stand-off distance became uncommonly large; in another case the shock itself oscillated between a necked and a smooth shock. These two cases have only been observed and have not been analyzed as of this writing.

The results of these tests indicate that a cylindrical afterbody will indeed turn the capsule entry body around to zero angle of attack but the length required to do this may be prohibitive. The results further indicate the internal configuration of the cylindrical afterbodies is not an important factor.

5. Optimum Supersonic Ejector Performance Characteristics, *R. D. Wood and R. J. Hemstead*

The principle of the jet ejector has been known and employed by industry for many years. The simplicity of its construction, low cost, light weight, and lack of moving parts has attracted many diverse applications. Although widely used, engineering design has usually been based on empirical techniques. Recent attempts to extend the use of the jet pump to a host of potential new applications are subject to question because of a lack of understanding of even the general performance characteristics of a typical configuration.

Comprehension of the fundamental fluid mechanics problem of the entrainment of a secondary stream by means of a high-velocity jet is handicapped by lack of knowledge of the turbulent mixing process and made even more complex by the proximity of the walls of the mixing tube. Recent applications of the ejector principle, as reported, may be found in the field of propulsion where air-augmented rockets or ejector ramjets are attracting interest because of the possibility of improved thrust performance with present launch systems by means of an ejector shroud (Ref. 6). In the field of space science the ejector has been advanced as a simple collection apparatus of aerosol matter for exobiological life detection experiments (*SPS 37-29*, Vol. IV p. 119, and Ref. 7). Extensive experimental and theoretical studies of zero secondary flow ejectors have been carried out as a means of altitude simulation of rocket nozzles (Refs. 8 and 9). High-performance ejector systems are also frequently used as jet pumps at the downstream end of low-density and hypersonic wind tunnels to reduce minimum starting and running compression ratios required for operation (Refs. 10 and 11).

a. Basic equations. Because there is a lack of fundamental understanding of the actual entrainment process in the above experiments, resort is frequently made to simplifying assumptions to obtain at least a partial solution to the problem. Such simplification invariably has the effect of limited application to a narrow range of conditions and geometric configurations which the ejector can assume. Now, a general understanding of the ejector performance characteristics can be obtained only by means of a simple one-dimensional compressible flow analysis.

The basic analysis utilizing this technique was first given by Kiselev (Ref. 12), extended by Wood (Ref. 13) and later by Fabri and Siestrunck (Ref. 14). Application of the conservation equations of mass, momentum, and energy produces equations of the form below in order to describe the configuration shown in Fig. 12 (Ref. 13).
Continuity equation (1-3)

$$\frac{P_{03}}{P_3}\,\theta(M_3^*) = \left(1 + \frac{w_2}{w_1}\right)(T_{03}/T_{01})^{1/2}\,\frac{A_1}{A_3}\,\theta(M_1^*)\,\frac{P_{01}}{P_3}. \tag{1}$$

Momentum equation (1, 2–3)

$$\frac{W_2}{W_1}(T_{02}/T_{01})^{1/2}\left(M_2^* + \frac{1}{M_2^*}\right) + \left(M_1^* + \frac{1}{M_1^*}\right) = \left(1 + \frac{W_2}{W_1}\right)(T_{03}/T_{01})^{1/2}\left(M_3^* + \frac{1}{M_3^*}\right) \tag{2}$$

Continuity equation (1-2)

$$\frac{P_{02}}{P_3} = \frac{W_2}{W_1}(T_{02}/T_{01})^{1/2}\,\frac{A_1}{A_2}\,\frac{\theta(M_1^*)}{\theta(M_2^*)}\,\frac{P_{01}}{P_3}. \tag{3}$$

Energy equation (1, 2–3)

$$\frac{T_{03}}{T_{01}} = \frac{1 + \dfrac{W_2}{W_1}\dfrac{T_{02}}{T_{01}}}{\left(1 + \dfrac{W_2}{W_1}\right)}. \tag{4}$$

The flow process is considered, at this point, to be inviscid and the specific heat ratio γ constant. These equations

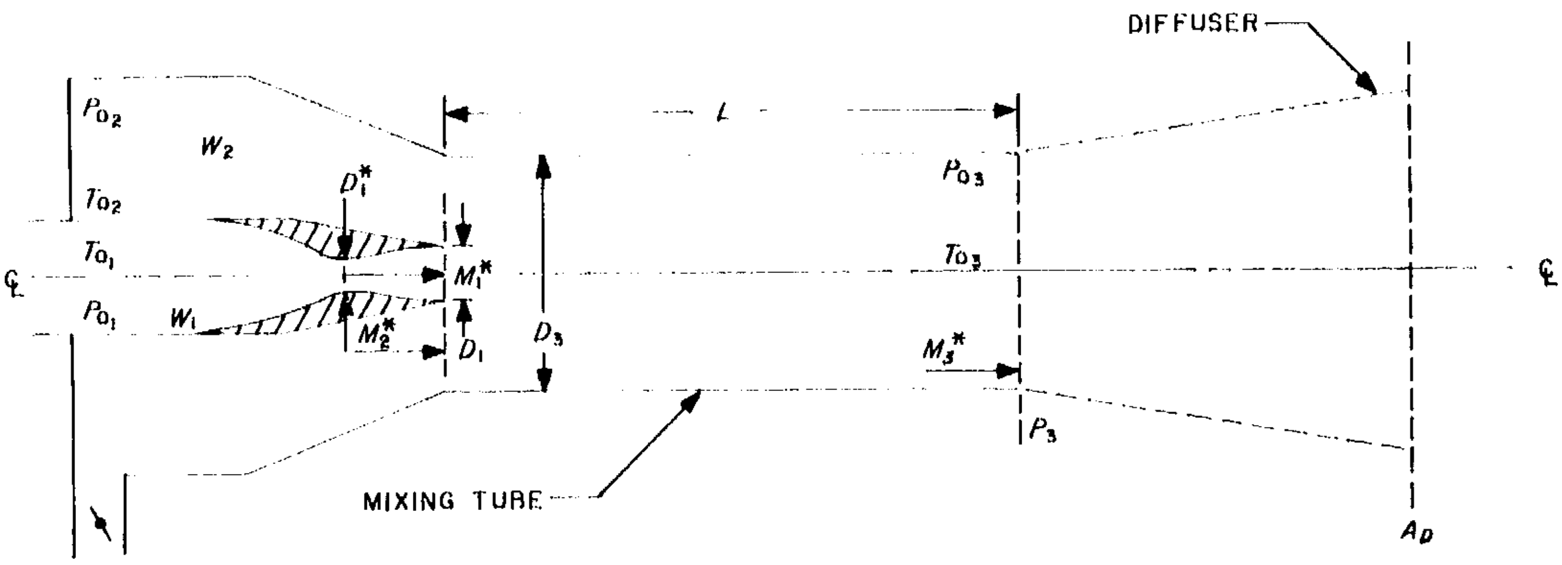


Fig. 12. Ejector schematic

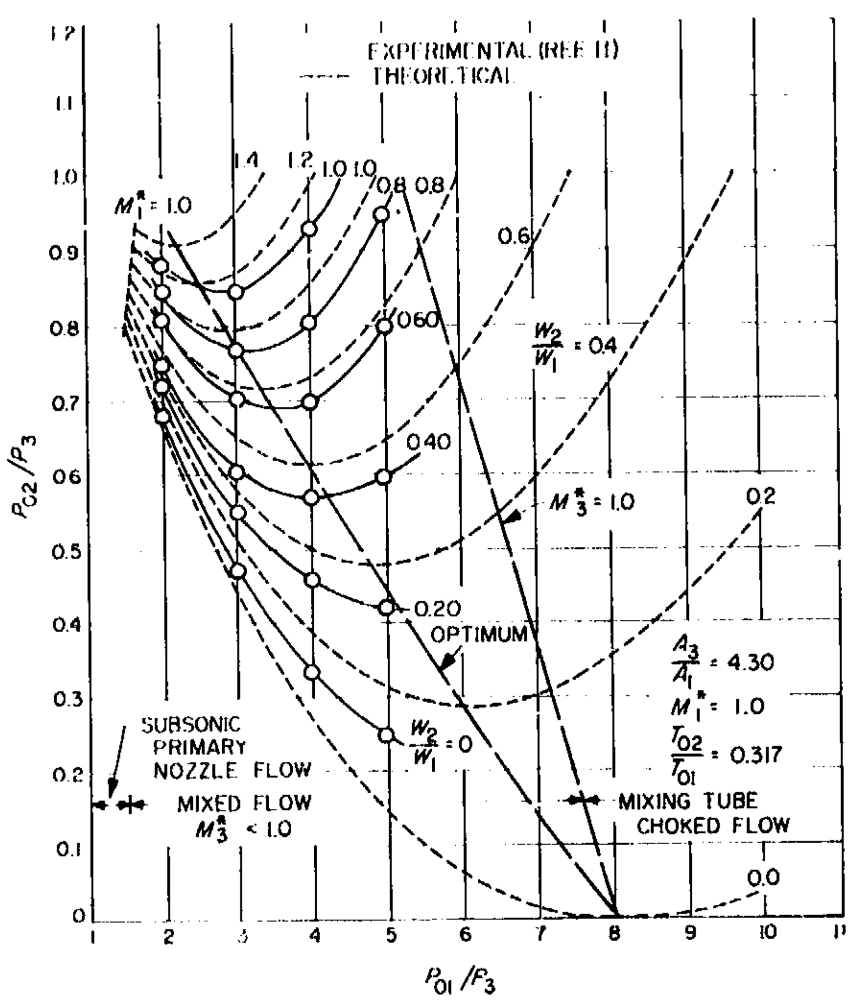


Fig. 13. Typical ejector performance with sonic nozzle ($M_1^* = 1.0$)

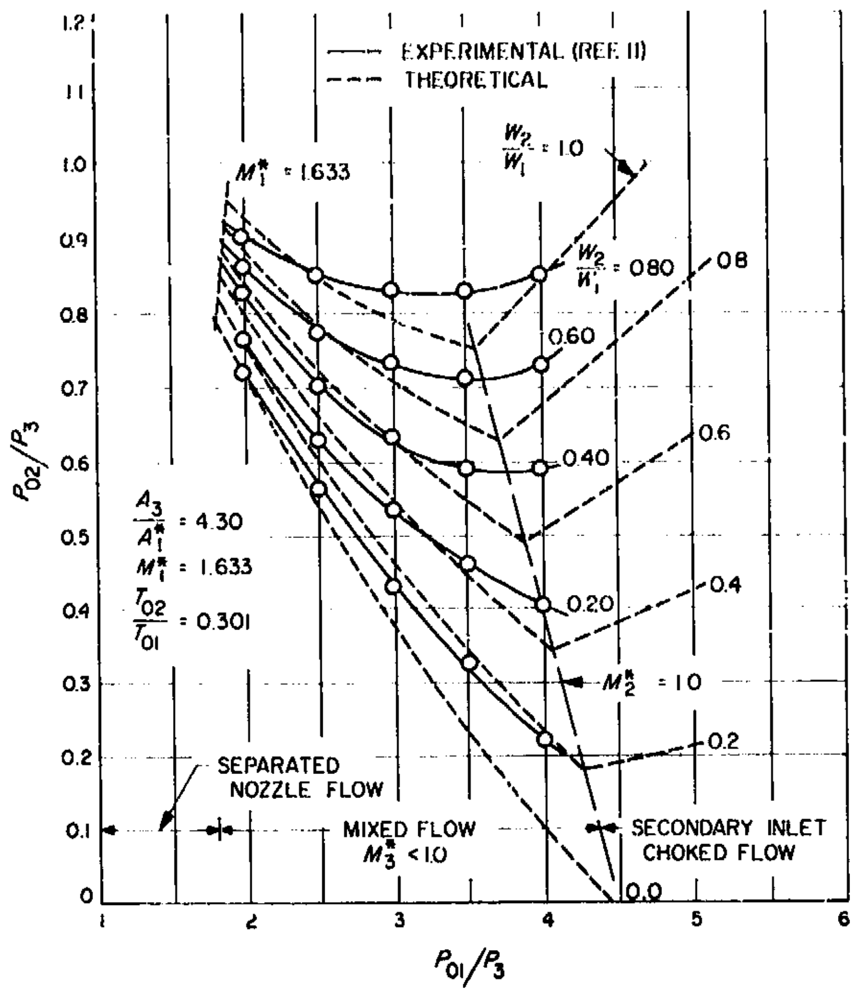


Fig. 14. Typical ejector performance with supersonic nozzle ($M_1^* = 1.633$)

are expressed, for simplicity, in terms of the dimensionless velocity M^* which is related to the Mach number by

$$M^2 = \frac{\frac{2}{\gamma+1} M^{*2}}{\left[1 - \frac{\gamma-1}{\gamma+1} M^{*2}\right]} \tag{5}$$

and the mass flux function by

$$\theta(M^*) = \frac{A^*}{A} = \left(\frac{\gamma+1}{2}\right)^{1/\gamma-1} M^*\left[1 - \frac{\gamma-1}{\gamma+1} M^{*2}\right]^{1/\gamma-1} \tag{6}$$

Solution to Eqs. (1) to (6), for a particular geometric configuration, is found by selection of values of

M_1^*, T_{02}/T_{01}, and W_2/W_1 and determining the change of P_{02}/P_3 for a variation in the primary jet pressure ratio P_{01}/P_3. These equations were programmed for the IBM 7094 and typical performance curves are shown in Figs. 13 and 14 for both a sonic and supersonic ($M = 2.0$) primary nozzle. A limited amount of experimental data is shown for comparison and it is noted that the simple theory correctly describes the general trend of the data. The difference between theory and experiment can be attributed to viscous mixing losses and simple shear stress losses along the mixing tube wall (Ref. 14).

b. Optimum performance. The particular problem of interest in this report is the determination and significance of the "optimum" point of the ejector performance curve or the condition at which, for a given mass flow ratio W_2/W_1, the minimum secondary pressure ratio P_{02}/P_3 is obtained. This is normally the condition desired in the actual operation of any real configuration. The "optimum" point of ejector performance was found by the following method:

From Eq. (3)

$$\frac{d(P_{02}/P_3)}{d(P_{01}/P_3)} = \frac{[\theta(M_2^*) - P_{01}/P_3\, d\theta(M_2^*)/d(P_{01}/P_3)]\, \frac{W_2}{W_1}\, (T_{02}/T_{01})^{1/2}\, \frac{A_1}{A_2}\theta(M_1^*)}{[\theta(M_2^*)]^2}. \tag{7}$$

By setting this derivative equal to zero we obtain

$$\frac{d\theta(M_2^*)}{d(P_{01}/P_3)} = \frac{\theta(M_2^*)}{P_{01}/P_3}. \tag{8}$$

But by the chain rule we write

$$\frac{d\theta(M_2^*)}{d(P_{01}/P_3)} = \frac{d\theta(M_2^*)}{dM_2^*} \times \frac{dM_2^*}{dM_3^*} \times \frac{dM_3^*}{d(P_{01}/P_3)}. \tag{9}$$

Then from Eq. (8)

$$\frac{d\theta(M_2^*)}{dM_2^*} = \left(\frac{\gamma+1}{\gamma}\right)\left[1 - \frac{\gamma-1}{\gamma+1}M_2^{*2}\right]^{-\gamma/\gamma-1}$$
$$\times \left[1 - \frac{\gamma-1}{\gamma+1}M_2^{*2}\left(1 + \frac{2}{\gamma-1}\right)\right] \tag{10}$$

and from Eq. (2)

$$\frac{dM_2^*}{dM_3^*} = \frac{\left(1 + \frac{W_2}{W_1}\right)(T_{03}/T_{01})^{1/2}\left[1 - \frac{1}{M_3^{*2}}\right]}{\frac{W_2}{W_1}(T_{02}/T_{01})^{1/2}\left[1 - \frac{1}{M_2^{*2}}\right]} \tag{11}$$

and from Eq. (1) where

$$\frac{P_{03}}{P_3}\theta(M_3^*) = \frac{\left(\frac{\gamma+1}{2}\right)^{1/\gamma-1} M_3^*}{\left[1 - \frac{\gamma-1}{\gamma+1}M_3^{*2}\right]} \tag{12}$$

we obtain

$$\frac{dM_3^*}{d(P_{01}/P_3)} = \frac{\left(1 + \frac{W_2}{W_1}\right)(T_{03}/T_{01})^{1/2}\frac{A_1}{A_3}\theta(M_1^*)}{\left(\frac{\gamma+1}{2}\right)^{1/\gamma-1}}$$
$$\times \frac{\left[1 - \frac{\gamma-1}{\gamma+1}M_3^{*2}\right]^2}{\left[1 + \frac{\gamma-1}{\gamma+1}M_3^{*2}\right]} \tag{13}$$

Substituting the results from Eqs. (10), (11), and (13) into Eq. (9) and equating the expressions for $d\theta(M_2^*)/d(P_{01}/P_3)$ in Eqs. (8) and (9), we obtain after considerable simplification

$$\frac{W_2}{W_1}(T_{02}/T_{01})^{1/2}\left[M_2^* - \frac{1}{M_2^*}\right]\left[1 - \frac{\gamma-1}{\gamma+1}M_2^{*2}\right]$$
$$\times\left[1 + \frac{\gamma+1}{\gamma+1}M_3^{*2}\right] - \left(1 + \frac{W_2}{W_1}\right)(T_{03}/T_{01})^{1/2}$$
$$\times\left[M_3^* - \frac{1}{M_3^*}\right]\left[1 - \frac{\gamma-1}{\gamma+1}M_3^{*2}\right]$$
$$\times\left[1 - \frac{\gamma-1}{\gamma+1}M_2^{*2}\left(1 + \frac{2}{\gamma-1}\right)\right] = 0. \tag{14}$$

If we consider simultaneously the above equation and Eq. (2), we have two equations in the variables M_2^* and M_3^*. If we stipulate from geometric considerations that M_2^* must lie between 0 and 1.0, Eq. (2) can be solved quadratically for M_2^*. Eq. (14) can then be simplified to an equation in only the unknown M_3^*. A short IBM 7094 program was written to solve these equations and solutions obtained for several primary jet nozzle Mach numbers.

*c. **Results.*** Figs. 15 and 16 show the results of these calculations for a sonic primary nozzle ($M_1^* = 1.0$) and for a supersonic nozzle with $M_1^* = 1.633$ ($M = 2.0$). A significant difference in the limiting conditions is apparent between the two cases. It also may be noted from these equations that the optimum velocity combination is independent of the geometric (A_2/A_1) configuration.

A fundamental point to consider is that the condition of maximum secondary compression ratio (P_3/P_{02}) occurs with subsonic flow in the mixing tube or $M_2^* \leq 1.0$ irrespective of primary nozzle velocity. This particular regime has been termed a "mixed flow pattern" by Fabri and Siestrunck (Ref. 14). In the case of the primary supersonic nozzle, performance may be limited by choking ($M_2^* = 1.0$) of the secondary flow at the mixing tube entrance. For values of the mass flow ratio $W_2/W_1 \rightarrow 0$ the simple one dimensional approach fails since $M_2^* \rightarrow 0$.

Figs. 15 and 16 also indicate the effect of the use of a heated primary jet or $T_{02}/T_{01} < 1.0$ on optimum ejector performance. It is seen that one effect of heating is to reduce the value of M_2^* or to delay the onset of secondary flow choking. Actual performance increases of P_{02}/P_3 are proportional to $[(T_{02}/T_{01})^{1/2}/\theta\,(M_2^*)]$. The frequent use of the approximate similarity relationship $W_2/W_1\,(T_{02}/T_{01})^{1/2}$ to estimate temperature effects obviously can be applied only for values of $M_2^* \rightarrow 1.0$ or for very low mass flow ratios W_2/W_1.

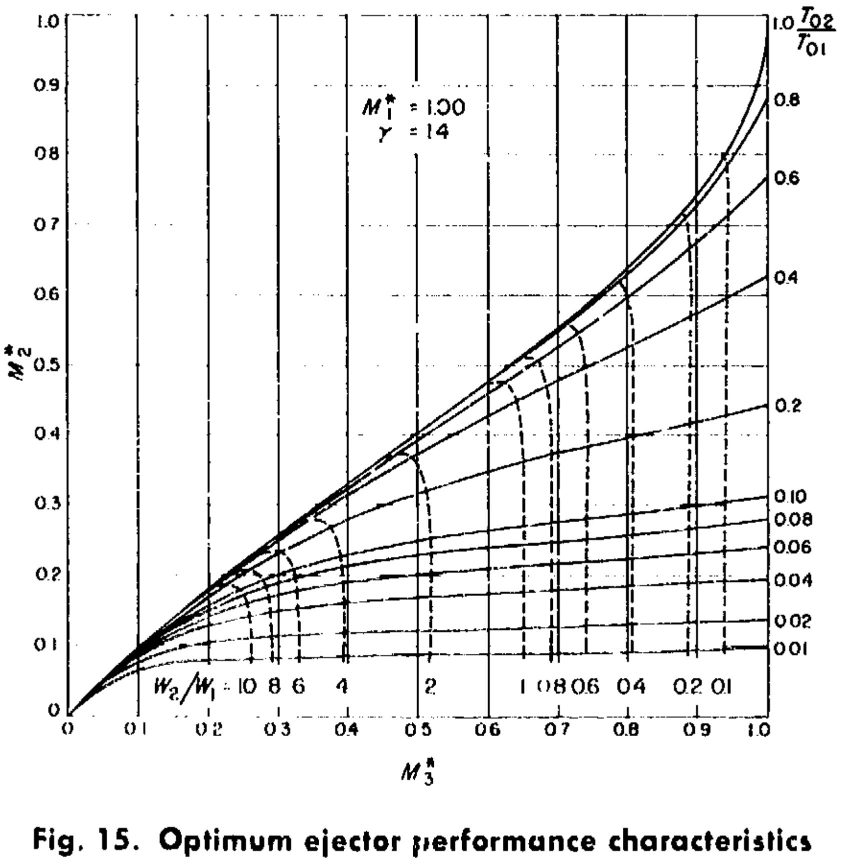


Fig. 15. Optimum ejector performance characteristics with sonic nozzle ($M_1^* = 1.0$)

*d. **Geometric configuration.*** Only the simplest constant area ejector configurations have been considered in this report. Obviously the choice of a simple center primary jet could be modified or an annular jet could be considered. A wide variety of geometric diffusers could be attached to the mixing tube exit or the discharge pressure controlled by means of a vacuum pump. Although a subsequent report will consider the use of a divergent and convergent–divergent diffuser it is evident that "optimum" ejector performance will continue to occur at the subsonic mixing tube velocities determined by the preceding analysis. The actual values of W_2/W_1, P_{02}/P_3, P_{01}/P_3 will change to satisfy geometric and mixing tube exit conditions.

*e. **Viscous effects.*** Numerous attempts have been made to develop simple empirical methods to obtain better correlation between theory and experimental results. As mentioned previously, a lack of understanding of the actual mixing and entrainment process limits the applicability of a specific method. Fabri and Siestrunck were able to obtain correlation with their experimental data by means of a simple wall shearing stress pressure loss calculation, but their theoretical calculations appear to be incorrect (Ref. 14). Application of a constant-pressure

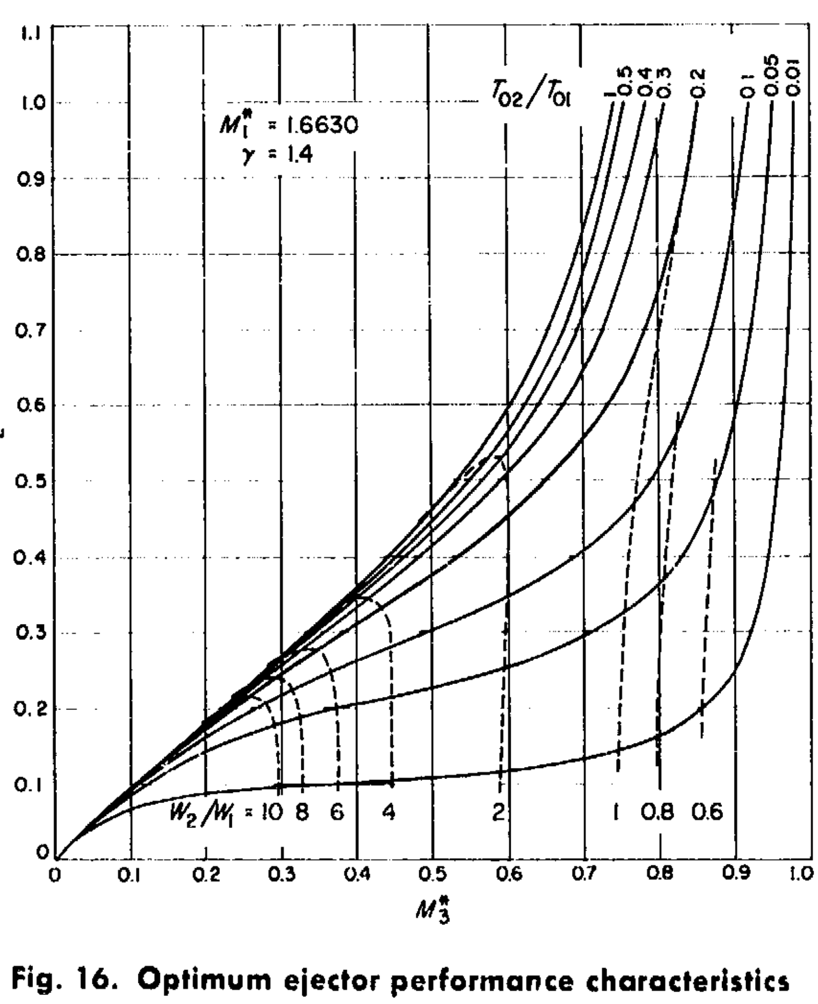


Fig. 16. Optimum ejector performance characteristics with supersonic nozzle ($M_1^* = 1.633$)

turbulent mixing process by Chow and Addy produced correlation with their own limited experimental data (Ref. 15). German, Bauer, and Panesci were able to obtain fair correlation with extensive experimental data under zero secondary flow conditions by means of two-dimensional supersonic base pressure theory modified by wall recompression analyses (Ref. 9). In a rather unusual test conducted in the JPL 21-in. Hypersonic Wind Tunnel, Hill and Nicholson were able to develop an entrainment coefficient of the mixing layer by measurement of the rate of fluid entrainment from the "dead-water" region of a spike used to create a conical region of separated flow ahead of a flat-faced cylindrical model (Ref. 16).

Analysis of simple axisymmetric base pressure flow indicates this basic flow mechanism: for equilibrium base pressure, the mass flow entrained from the stagnation region by the mixing process must equal the mass flow reversed into the region by the pressure rise through the recompression area (Ref. 9). Employing a recirculation concept based on this flow mechanism, the author is attempting to develop a correlation coefficient. A comparison of the various techniques discussed here is under way and will be reported on later.

B. Hypervelocity Laboratory

T. L. Babineau, B. R. Riale, W. A. Menard, and G. M. Thomas

1. Preliminary Performance of a Free Piston Shock Tube Driver, *T. L. Babineaux and B. R. Riale*

A scaled-up version of the free piston shock tube now in use at Harvard University has been constructed and is now being evaluated in the hypervelocity laboratory of the Experimental Entry Group. The free piston compression technique is depicted schematically in Fig. 17.

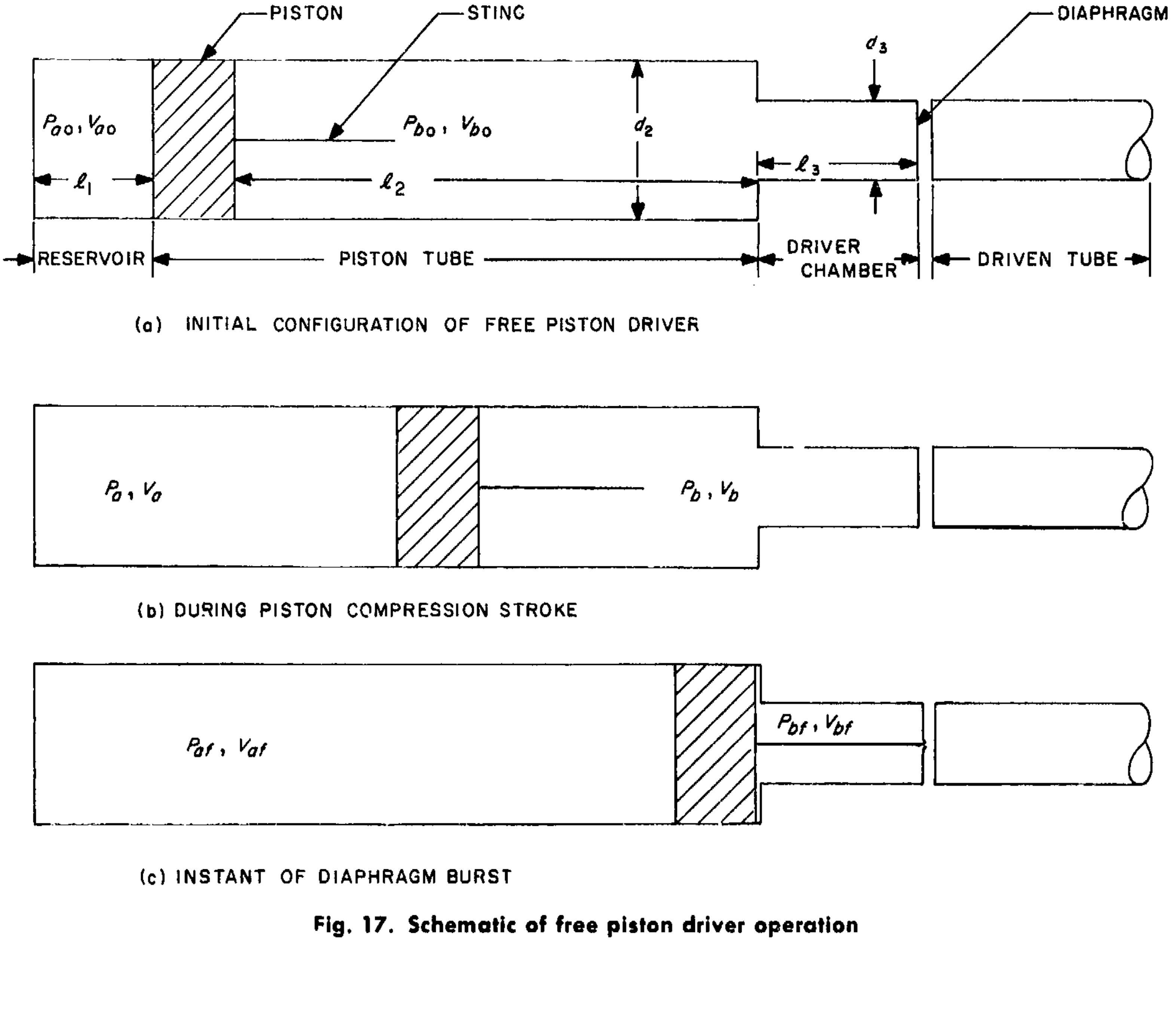


Fig. 17. Schematic of free piston driver operation

The apparatus is divided into three sections: the reservoir, the piston tube, and the shock tube driver chamber. The piston tube driver chamber combination is filled with the shock tube driver gas (helium) at low pressure. The reservoir is filled with the piston driving gas (helium or nitrogen) at high pressure while the piston is held at the rear of the piston tube by a mechanical holder and release device. When the piston is released, it accelerates forward, compressing the driver gas ahead of it. The initial pressure ratio across the piston can be adjusted to a value which will allow the piston to just reach the rear of the driver chamber before the force of the compressed gas reverses its direction. At that instant a sting, mounted on the front face of the piston, pierces the diaphragm and initiates the shock tube flow. A complete theoretical analysis of the piston motion may be found in Ref. 17.

Prior to actual design, a theoretical study (Ref. 17) was conducted to determine the parameters which affected the compression efficiency. The design parameters evolving as a result of this study and other considerations are shown in Table 3. The free piston shock tube is shown in Fig. 18. To evaluate the performance

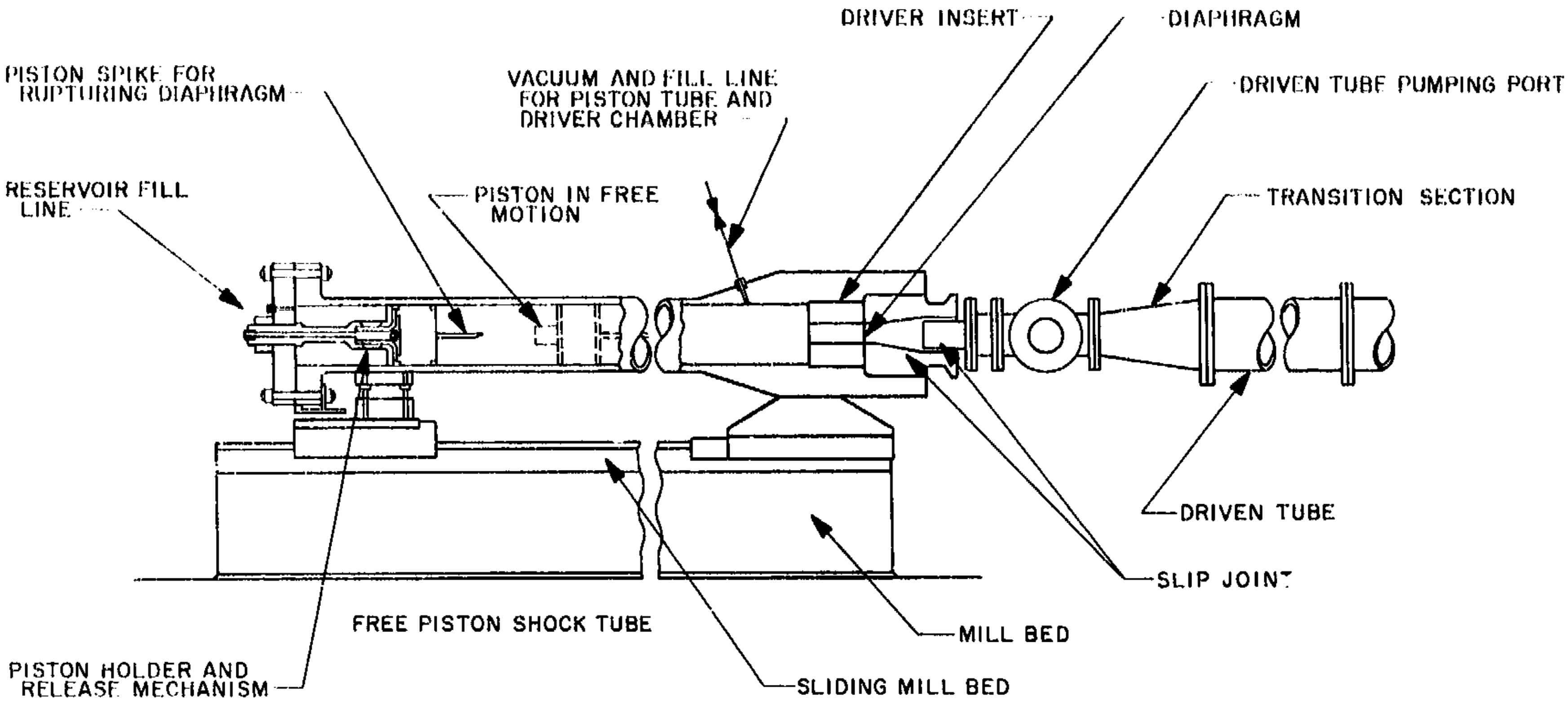


Fig. 18. Free piston shock tube

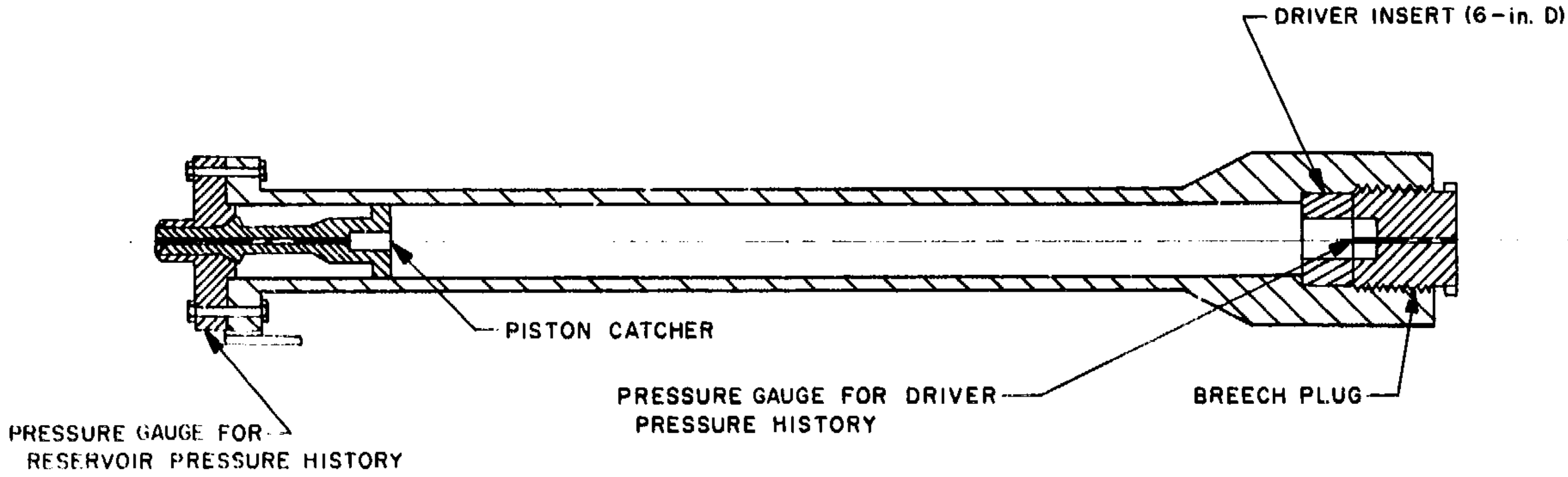


Fig. 19. Test configuration for driver performance evaluation

of the driver, a breech plug which would replace the diaphragm was constructed (Fig. 19). Pressure transducers were installed to measure the pressure history of the gas driving the piston and of the gas being compressed. Along with these measurements, final volume measurements were made and the polytropic exponent characterizing the thermodynamics of the process are determined from the relation

$$\frac{P_{b0}}{P_{bf}} = \left(\frac{V_{bf}}{V_{b0}}\right)^n$$

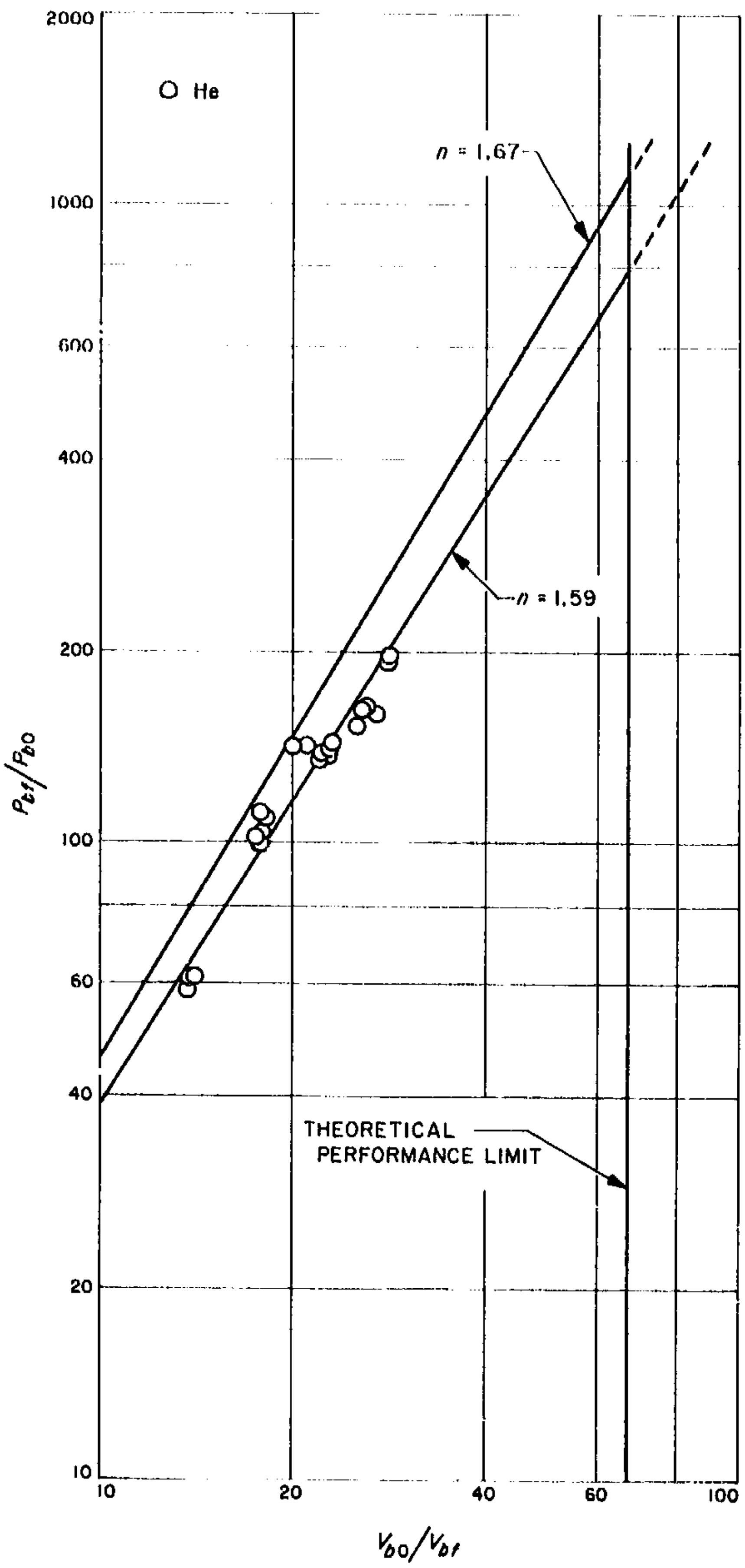


Fig. 20. Experimental results compared to ideal ($\gamma = 5/3$) from the relation $P_{b0}/P_{bf} = (V_{bf}/V_{b0})^n$

Table 3. Design parameters

Parameter	Dimension, in.	Nomenclature
l_1	25.5	Reservoir length
l_2	136.375	Piston tube length
d_2	12.00	Piston tube diameter
d_3	6.00	Driver chamber diameter
	4.50	
	3.00	
l_3	8.248	Driver chamber length (with breech plug)
	8.750	Driver chamber length (with diaphragm)
R		$l_3 d_3^2 / l_2 d_2^2$
$(V_{b0}/V_{bf})_{max}$		Maximum volume ratio
λ		l_1/l_2

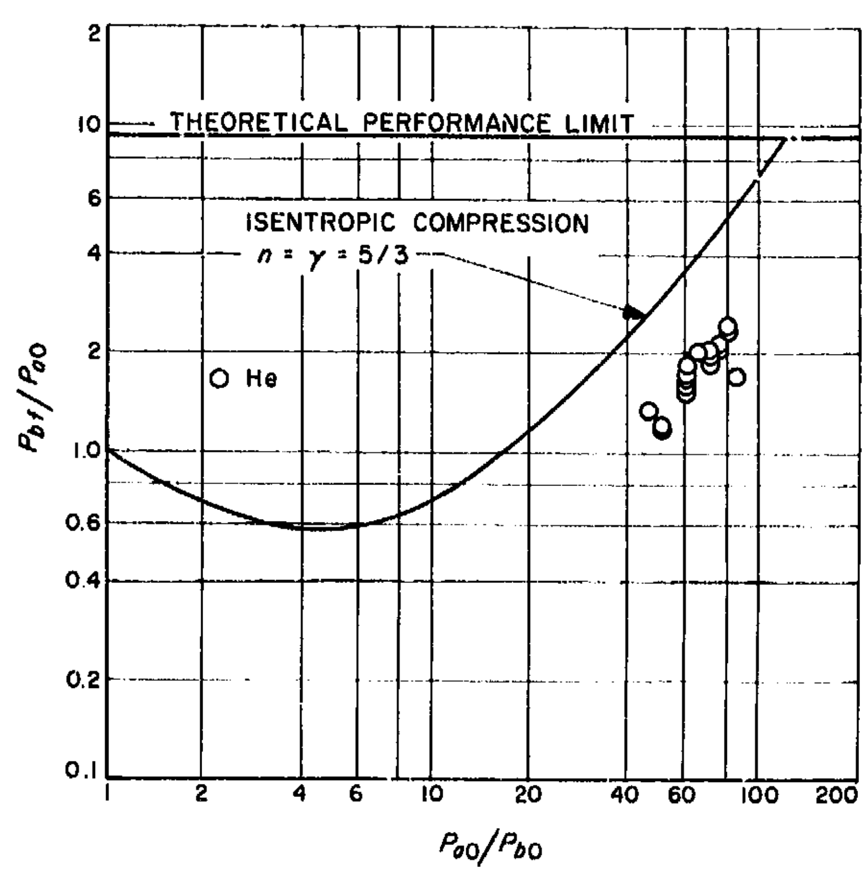


Fig. 21. Ratio of final volume to initial pressure in the piston tube reservoir as a function of pressure ratio across the piston

A series of shots was planned to determine the following: (1) the final volume and pressure as a function of the initial pressure ratio across the piston, P_{a0}/P_{b0}; (2) the variation of losses with initial pressure in the piston tube, P_{b0}; (3) variation of losses using a light (115 lb) and heavy (313 lb) piston; and (4) variation of losses as a function of the driver chamber volume. Some results of Phase 1 are presented in Figs. 20, 21, and 22.

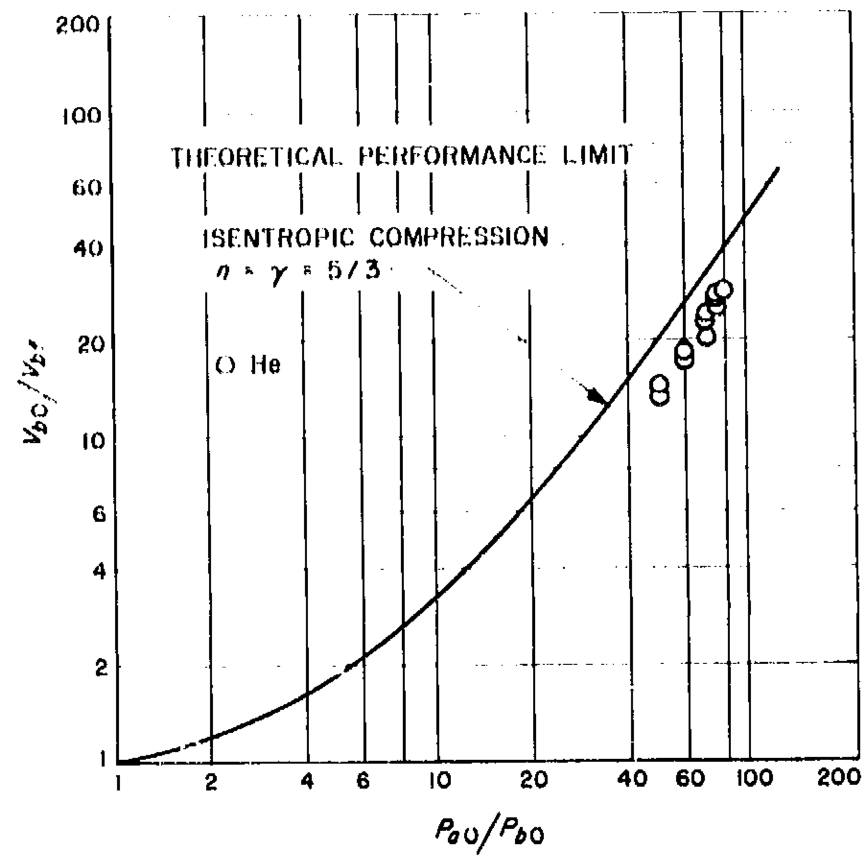


Fig. 22. Volume ratio obtainable as the pressure ratio across piston is increased

2. Progress with Spectrograph Calibration Experiments, *W. A. Menard and G. M. Thomas*

A Jarrell-Ash F/6.3 plane grating spectrograph has been modified for shock tube use. When installed, it will have the capability of measuring six channels of spectral radiation data from a single shock tube run. Lower wavelength coverage than obtainable with existing instrumentation (Ref. 18 and *SPS 37-26*, Vol. IV, p. 85) is also expected.

The dispersion of the spectrum at the focal plane has been measured accurately from spectrograms of several standard light sources. In Order I, linear dispersion of 9.1 Å/mm at 5000 Å to 10.0 Å/mm at 3000 Å was observed. At 3500 Å in Order II the dispersion was found to be 3.9 Å/mm, nearly that predicted by theory.

Fig. 23 illustrates the resolving power of the spectrograph. The spectrograms are of the dominant CN bands produced in the plasma of a carbon arc. The background luminosity is radiation from the positive electrode. Orders I and II were used to obtain the top and bottom spectrograms, respectively. The band head vibrational transitions are identified, along with their respective wavelengths. Much of the rotational line structure is seen in the spectrograms, particularly in Order II.

A problem that always exists in grating instruments is the effect of overlapping orders. What this means is that in some region the spectrum of Order II overlaps the spectrum of Order I. A detector in this region would measure the sum of radiation of wavelength λ from Order I and wavelength $\lambda/2$ from Order II. To separate the orders, the generally accepted procedure is to use appropriate filters. As an example, if one is working in Order I with λ not greater than 6000 Å, then a filter having zero transmission below 3000 Å must be used to eliminate the overlapping radiation from Order II.

Gratings are blazed to reflect maximum energy into a given angle with respect to the surface. The angle is proportional to the product $m\lambda$, where m is the order number. Our grating is blazed for 5000 Å in Order I, thus on both sides of this wavelength the grating efficiency drops off. In fact it has been found that it is nearly impossible to make measurements below 2600 Å with the present detector. However, since the angle of maximum efficiency is proportional to $m\lambda$, one immediately looks to Order II as a possible place to investigate the region around 2500 Å. This has been done by looking at the 2536-Å Hg line in both orders. It was found to have 20 times more intensity in Order II! To work between 2000 and 3000 Å in Order II, a filter must be used that has good transmission in this region but cuts out the 4000 to 6000 Å overlapping Order I spectrum.

An over-all efficiency factor for the spectrograph has been obtained by assuming $V \sim RFIE$. V is the photomultiplier output voltage, R is the photomultiplier response, F is the total transmission of any filters in the optical system, I is the intensity of the calibration source, and E is the desired efficiency. For any two wavelengths the ratio of efficiency factors is

$$\frac{E_2}{E_1} = \frac{V_2}{V_1}\left(\frac{R_1}{R_2}\frac{F_1}{F_2}\frac{I_1}{I_2}\right).$$

R was available from the manufacturer's manual, F was measured, and I is known over most of the spectrum for the carbon arc source. By measuring V it was, thus,

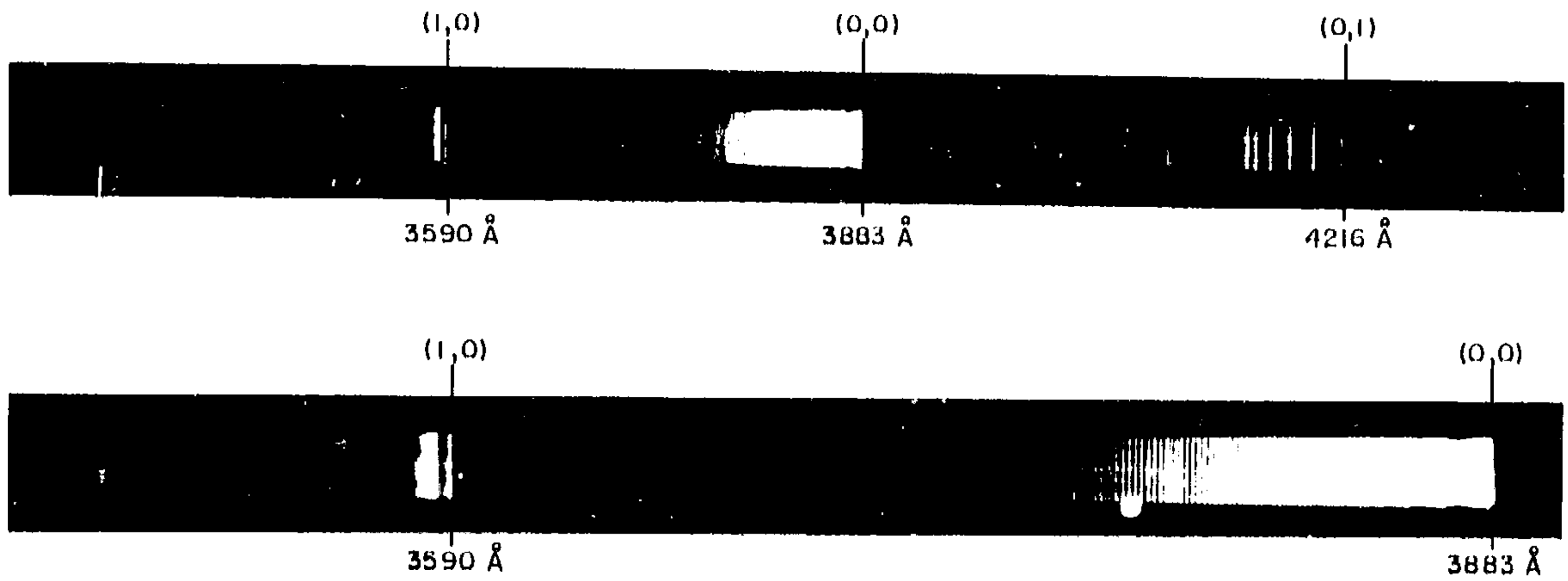


Fig. 23. Spectrograms of CN bands from the plasma of a carbon arc

possible to compute the efficiency of the spectrograph. The efficiency factor was found to be a smooth function of wavelength and has the same general distribution as the grating efficiency.

One effect observed is that visible stray light from the grating is reaching the detector and interfering with measurements below 2600 Å. The nature of the stray light is such that interference screening within the spectrograph does not solve the problem. Working in Order II does not help because the stray light is there also. Three approaches to solving the problem have been examined:

(1) A filter that has good transmission from 2000 to 3000 Å but cuts out the visible obviously would do the job. Unfortunately, filters that have adequate transmission much below 2400 Å are not available.

(2) Another method is to coat the window of the photomultiplier tube with sodium salicylate ($NaC_7H_5O_3$) which is often used in ultraviolet work because of its high and relatively constant quantum yield characteristics. In the usual application a thin layer (2–4 mg/cm^2) is applied to the photomultiplier tube window (which does not transmit ultraviolet radiation). As incident ultraviolet radiation passes through the coating, molecules are excited, and as they relax back to their ground state, they emit photons of visible wavelength. The visible light produced is then transmitted through the photomultiplier window to the sensing element. Knowing the quantum efficiency of the sodium salicylate, it is possible to calibrate the coated photomultiplier tube. For a thin coating of this type, visible as well as ultraviolet radiation is transmitted through the layer. A thick coat should attenuate, by scattering, the specular transmission of stray light more than it does the fluorescent light, and so improve the stray light characteristics of the system (Ref. 19).

Coatings of various thicknesses were prepared by spraying a solution of sodium salicylate and methyl alcohol onto microscope slides with an artist's airbrush. The slides were placed in front of the photomultiplier with the coated surface facing the incident radiation. The photomultiplier signal was recorded for λ between 2000 and 3000 Å. It was found that only with the thickest coating (7 mg/cm^2) was there a noticeable reduction in stray light. The reduction was not sufficient for our purposes, however, and thicker layers are not practical because they become flaky and nonuniform. In principle, the technique looks good but a better coating process is needed to obtain thicker layers.

(3) A third method is to use a detector which is insensitive to visible light, yet highly sensitive to the ultraviolet. Detectors possessing this quality are termed "solar blind." Several types are being investigated.

References

1. Jaffe, P., and Prislin, R. H., "Effect of Boundary Layer Transition on Dynamic Stability over Large Amplitudes of Oscillation," *AIAA Preprint 64-427*, Presented at the First Annual Meeting, Washington, D. C., June 1964.

2. Jaffe, P. "Obtaining Free-Flight Dynamic Damping of an Axially Symmetric Body (At All Angles-of-Attack) in a Conventional Wind Tunnel," Technical Report No. 32-544, Jet Propulsion Laboratory, Pasadena, California, January 1964.

3. Prislin, R. H., "Captive Dynamic Stability Tests of Blunt Entry Configurations in the JPL Supersonic and Hypersonic Wind Tunnels," WT 20-548, Jet Propulsion Laboratory, Pasadena, California, May 1964.

4. Tobak, M., and Wehrend, W. R., "Stability Derivatives of Cones at Supersonic Speeds," NACA TN 3788, NASA, Washington, D. C., September 1956.

5. Sauerwein, H., "Application of the Piston Analogy to the Calculation of Stability Derivatives for Pointed Axially Symmetrical Bodies at High Mach Numbers," RAD-TM-61-40, October 1961.

6. "Marquardt Exploring Air Augmentation," *Missiles and Rockets*, August 17, 1964.

7. Quimby, F. H., "Concepts for Detection of Extraterrestrial Life," NASA SP-56, Washington, D.C., 1964.

8. Massier, P. F., and Roschke, E. J., "Application of Exhaust Diffusers for Rocket Engine Testing," Technical Memorandum No. 33-97, Jet Propulsion Laboratory, Pasadena, California, August 1962.

9. German, R. C., Bauer, R. C., and Panesci, J. H., "Methods for Determining the Performance of Ejector-Diffuser Systems Used in Rocket Altitude Simulation," AIAA Space Simulation Testing Conference, November 1964.

10. Howell, R. R., "Experimental Operating Performance of a Single-Stage Annular Air Ejector," NASA TN D-23, Washington, D.C., October 1959.

11. Hunczak, H. R., and Rousso, M. D., "Starting and Operating Limits of Two Supersonic Wind Tunnels Utilizing Auxiliary Air Injection Downstream of the Test Section," NACA TN 3262, NASA, Washington, D.C., 1954.

12. Kiselev, B. M., "Calculation of One-Dimensional Gas Flows," Brown University Technical Report No. F-TS-1209-IA (GDAM A9-7-27) (a translation from "Raschet Odnorazmernih Gasooih Techenic"), January 1949.

13. Wood, R. D., "Theoretical Ejector Performance and Comparison with Experimental Results," University of Minnesota, WADC Technical Report 54-556, Minneapolis, Minnesota, August 1954.

14. Fabri, J., Siestrunck, R., "Supersonic Air Ejectors," *Advances in Applied Mechanics*, Vol. V, 1958.

15. Chow, W. L., Addy, A. L., "Interaction Between Primary and Secondary Streams of Supersonic Ejector Systems and Their Performance Characteristics," *AIAA Journal*, Vol. 2, No. 4, April 1964.

References (Cont'd)

16. Hill, J. A. F., and Nicholson, J. E., "Compressibility Effects on Fluid Entrainment by Turbulent Mixing Layers," NASA Contractor Report CR-131, Washington, D.C., November 1964.

17. Roffe, G. A., "The Free Piston Shock Tube Driver: A Preliminary Theoretical Study," Technical Release No. 32-560, Jet Propulsion Laboratory, Pasadena, California, December 15, 1963.

18. Thomas, G. M., Menard, W. A., "Experimental Measurements of Nonequilibrium and Equilibrium Radiation from Planetary Atmospheres," AIAA Publication CP9, pp. 170–185, 1964.

19. Johnson, F. S., "Spectral Radiance of the Carbon Arc from 1900 to 2500 Angstroms," *Journal of the Optical Society of America*, Vol. 46, pp. 101–103, February 1956.

BLANK PAGE

PROPULSION DIVISION

XII. Solid Propellant Engineering

A. Low-Pressure Combustion

R. Sehgal and L. Strand

In earlier studies (Ref. 1) with polyurethane-type composite propellants, the results indicated that the extinction pressure, while independent of the burning geometry, was strongly dependent upon motor L^* and certain other parameters such as the Al concentration in the propellant. With increased concentration of Al in the propellant, the slope of the L^* versus extinction pressure relationship became steeper. In later studies (Ref. 2), when coarser aluminum was substituted in the propellant, the effect was to cause incomplete combustion at the low pressures, and the slope of the relationship and burning characteristics showed a trend approaching those of nonaluminized propellant. These results are shown in Figs. 1 and 2.

Additional studies have been made to determine the effect of oxidizer particle size on the L^* versus extinction pressure relationship. Two propellant formulations were used having the same percentage Al, ammonium perchlorate (AP), and binder in the propellant. One formulation had a unimodal AP distribution with an average particle size of approximately 400 μ (+48 mesh) and an average Al particle size of approximately 7 μ. The other formulation had a bimodal AP (70/30) distribution (unground, 400 μ; ground, 17 μ) with an Al particle size of approximately 31 μ. Burning rate data for these formulations in the low-pressure region were obtained by using the Crawford bomb strand burner. These data, along with JPL 540 propellant used in the earlier studies, are shown in Fig. 3. The formulation with the fine Al particle size, but with the coarser unimodal AP particle size distribution, suppressed the burning rate as expected, due to the reduced packing density. However, it was interesting to find that the other formulation with the coarser oxidizer particle size, but also with the coarser Al particle size, resulted in a higher burning rate than the JPL 540 propellant.

For low-pressure extinction tests, regressive burning 2.5-in. D and 4- and 4.5-in. long cylindrical charges were used in the 3-in. ID test motor. Using the standard squib-pellet igniter system, these modified propellants could

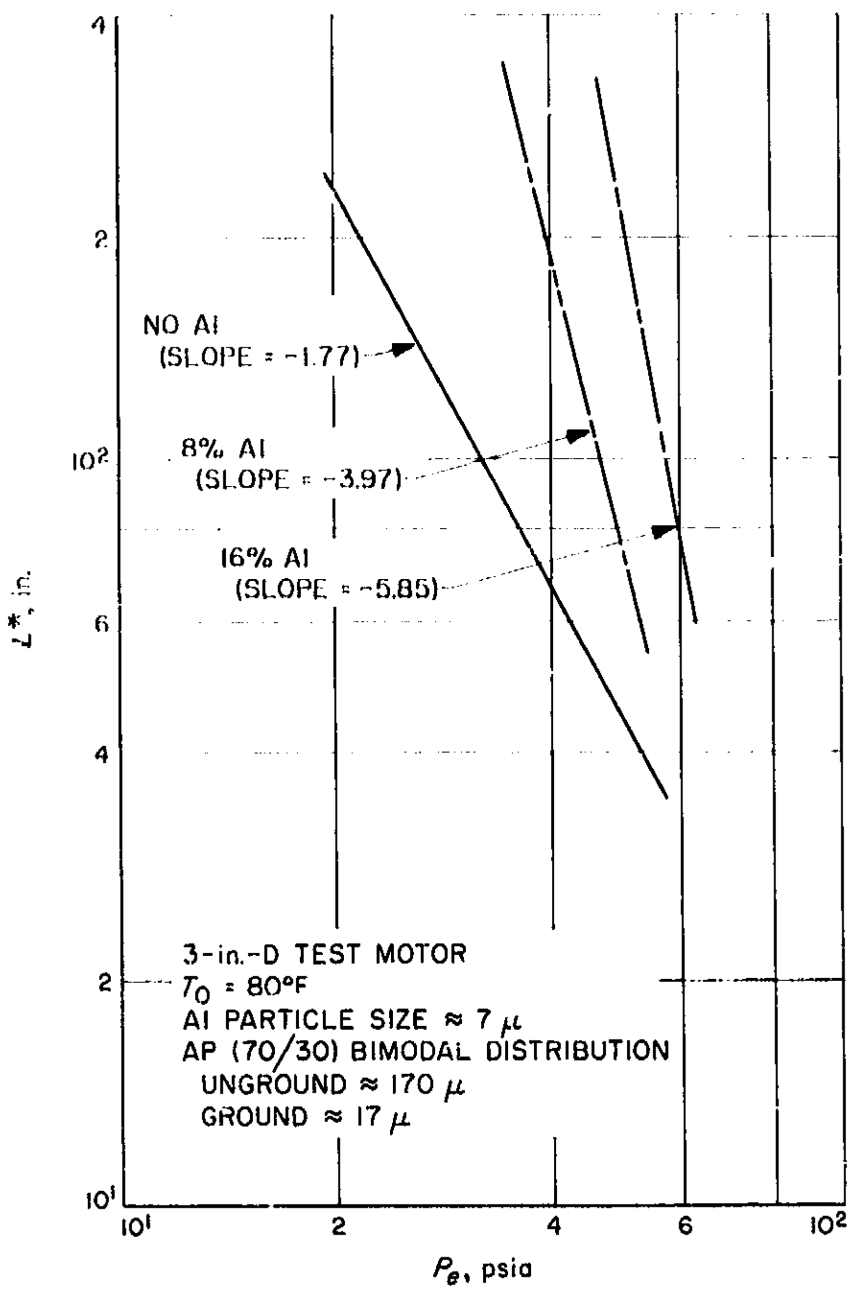


Fig. 1. Effect of Al concentration on L^* versus P_e relationship

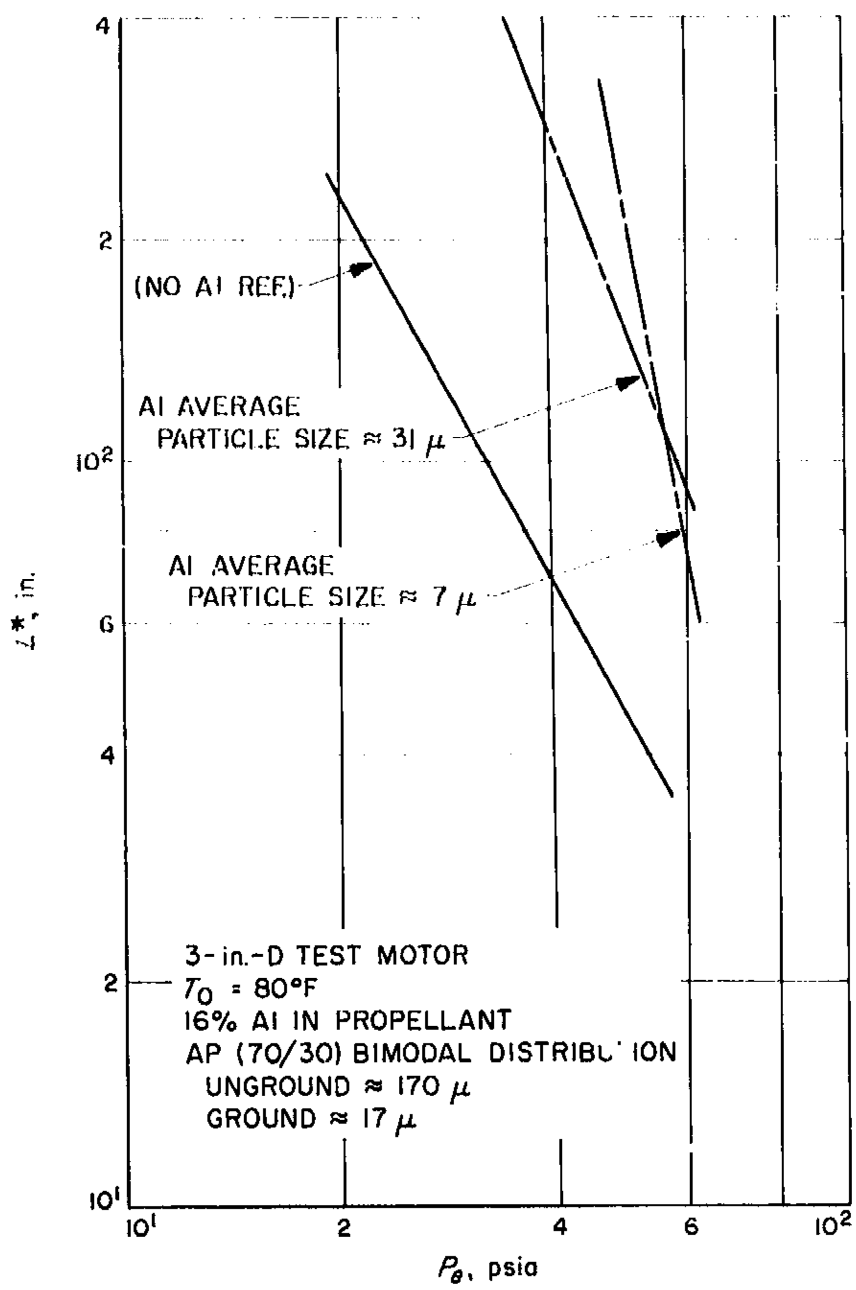


Fig. 2. Effect of Al particle size on L^* versus P_e relationship

not be ignited either under vacuum or atmospheric firing conditions. Ignition was not obtained under vacuum conditions even with a hotter igniter. Finally, satisfactory ignitions were obtained by attaching strips of a different easily ignitable propellant to these charges. The subject of sensitivity of ignition to pressure and the controlling factors for ignition with varying concentration and particle size of Al and AP in the propellant are currently under investigation. The results of the tests for the evaluation of coarser oxidizer particle size in propellant are shown in Figs. 4 and 5. The pertinent conclusion that can be drawn is that the slope of L^* versus extinction pressure relationship is not affected by the variation of oxidizer particle size in the propellant.

Another investigation was directed toward determining the lowest pressure ignitability limit in a Crawford bomb, of the various propellants used in this study. It was found that for the propellants which had the same percentage composition but different Al and AP particle size, the lowest pressure ignition limit was approximately 15 psia with a variation of less than 1 psia.

Some of the conclusions which can be drawn from the various phases of the experimental program conducted to date are briefly summarized as follows:

(1) For a given propellant the L^* versus extinction pressure relationship is a valid parameter for determining the low-pressure stable combustion limit.

(2) For nonaluminized propellants the slope of L^* versus extinction pressure relationship is $-2n$, as predicted theoretically for critical pressure.

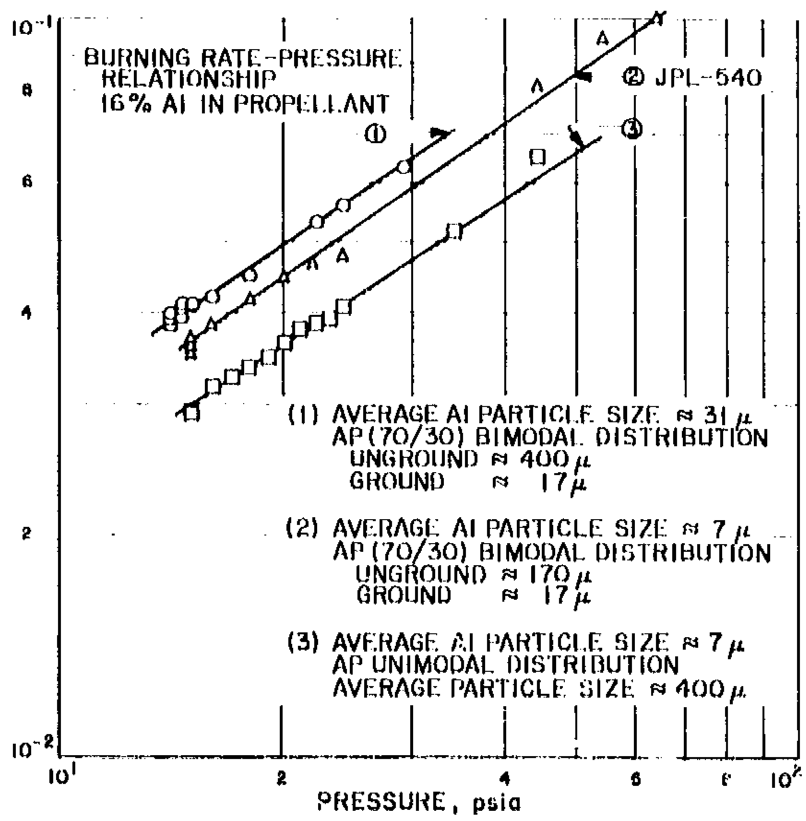


Fig. 3. Burning rate-pressure relationship of 16% Al in propellant

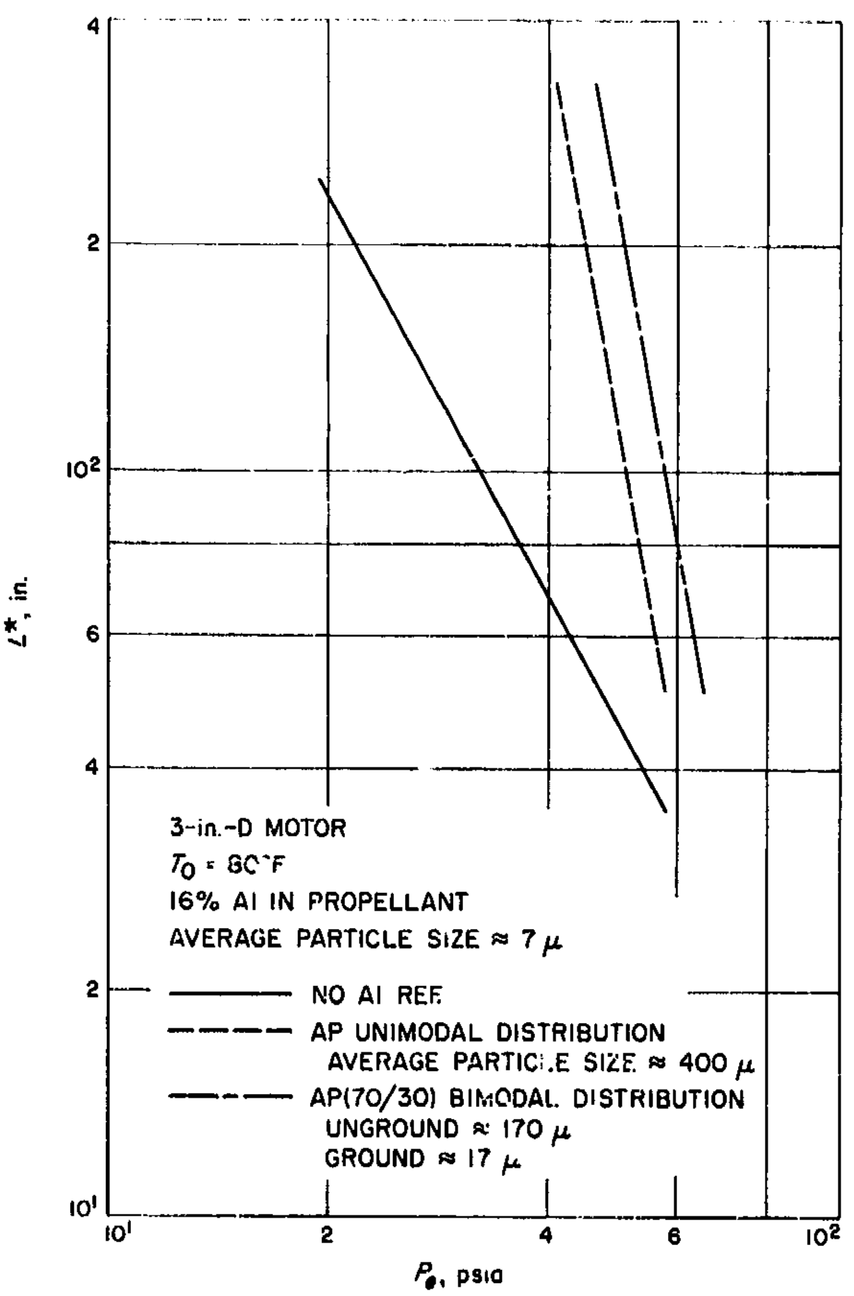


Fig. 4. Effect of AP particle size (bimodal distribution) on L^* versus P_c relationship

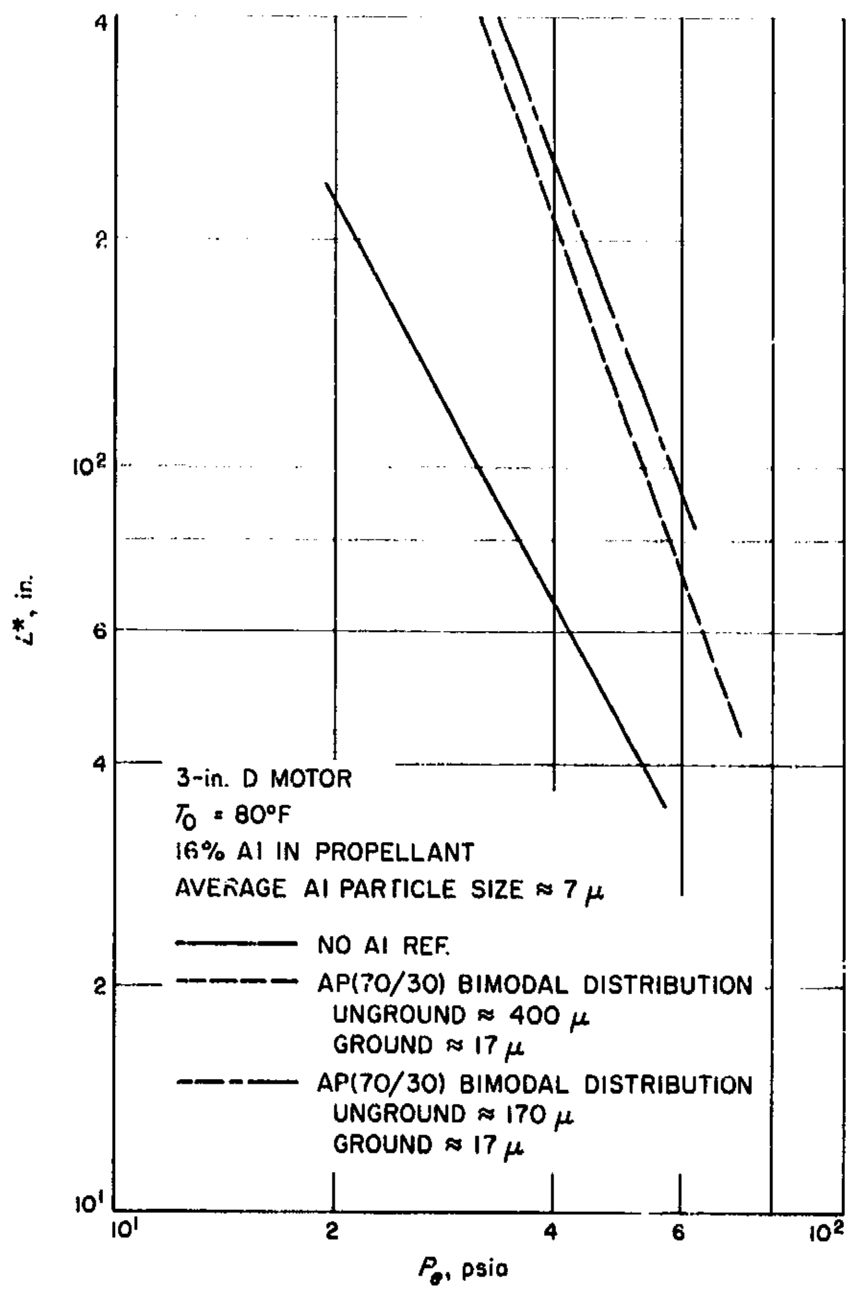


Fig. 5. Effect of AP particle size (unimodal distribution) on L^* versus P_c relationship

(3) The slope of the L^* versus extinction pressure relationship is affected by the presence of Al, the concentration of Al, and the Al particle size in propellant.

(4) The effect of coarser Al in propellant is to cause incomplete combustion at low pressures, and the burning and extinction characteristics tend to approach those of nonaluminized propellant.

(5) The variation in oxidizer particle size has a negligible effect on the L^* versus extinction pressure relationship. The small differences in extinction pressure at a given L^* are attributed to the variation in packing density.

B. Development of Sterilizable Pyrotechnic Devices

A. G. Benedict

1. Introduction

Electro-explosive devices, such as pinpullers, explosively actuated valves, explosive bolts, thrusters, explosively actuated switches, drogue-guns, and cable-cutters, are in common use in the aerospace field. Their main advantage over other comparable devices is in their compact energy-storage capability. Such devices are actuated by squib initiators, the output charges of the squibs being variously heat-producing (as for motor ignition), detonating (as for explosive bolts), gas-producing (as for cutters and actuators), and possibly light-producing for illumination or smoke-producing for signalling.

The designing of devices such as pinpullers to withstand thermal sterilization is generally rather routine, involving attention to details such as mechanical interference resulting from mismatches in temperature coefficients of expansion and the use of unsuitable organic materials, as in O-ring seals.

The problem of designing squib initiators to withstand thermal sterilization is more complex, mainly because materials suitable for squib construction are limited in range, thus limiting choice in avoidance of thermal mismatches. Also, the relatively small size of squibs precludes the use of sophisticated expansion joints dependent on precise machining.

However, there are several squib initiators on the market which would almost certainly withstand sterilization, and the question might reasonably be asked as to why a squib developmental program is needed. The answer lies in the fact that no known squib incorporates the range of features considered desirable for JPL spacecraft applications. These features include:

(1) Small size (and weight).

(2) Ability to withstand heat sterilization.

(3) Dual bridges.

(4) Ability to withstand a 1-w/1-amp no-fire.

(5) Exclusion of magnetic materials.

(6) Incorporation of an integral connector.

(7) Ability to withstand severe temperature cycling.

(8) Ability to withstand pressures of at least 30,000 psi, without significant venting or rupture.

(9) Ability to withstand electrostatic discharges of the order of 25 kv from a 500-$\mu\mu$f capacitor, applied between pins and case or between pins, under any pressure (altitude).

(10) Freedom from critical dependence on practically uninspectable qualities which, in turn, are critically dependent on workmanship.

With JPL's technical requirements in mind, a thorough review of presently available designs in the industry and those under development (such as the MSC standard *Apollo* initiator) was conducted. However, JPL's requirements differ so markedly from the designs reviewed that the decision was made to evolve a new basic initiator, similar to the *Apollo* initiator in that it can easily be adapted for heat production, gas production, or other purposes as required.

2. Design Approach

In the light of spacecraft requirements and the present state of initiator design, it was decided to evolve a squib design on a step-by-step basis on the premises that:

(1) The squib could be contained within a ⅜-in. 24-thread body.

(2) The mating connector would be similar to a small commercial connector, having four 0.04-in. diameter pins spaced 90 deg apart on a 0.160-in. base circle diameter, but weighing about ¼ as much as a comparable MIL-C-26482 typ .

The major steps envisaged in this approach were: (1) Selection and optimization of connector seals, (2) Selection and optimization of insulating insert, (3) Selection and optimization of bridge configuration, (4) Selection of match-head, (5) Optimization of body details, and (6) Optimization of mating connector.

It was considered that several time-honored restraints could be relaxed:

(1) Dielectric resistance between shorted pins and case need not be of the order of 2 MΩ (or more) as commonly specified; a resistance as low as 0.1 MΩ appears acceptable.

(2) Dielectric resistance between adjacent bridges need not be of the order of 2 MΩ (as typically specified when circuit isolation is required); a resistance of as low as 1,000 Ω appears acceptable.

(3) Hermetic (fusion) sealing is not required, provided venting is not excessive.

In order that the initiator could be used for ignition, gas production, or other special purposes, it appeared desirable to concurrently review the temperature degradation and other salient characteristics of possible explosive materials.

3. Squib Seal Materials

The most important problems involved in squib development are those associated with the seals.

Four types of material are used on present-day squib seals: epoxies, thermo-setting plastics, glass, and ceramic. The first two of these were discounted after preliminary considerations, mainly because no materials could be found with attractive combinations of temperature coefficients of expansion, thermal conductivity, and strength.

Glass seals can be manufactured reliably and economically, but only to withstand 30,000 psi; with glass seals, the problem of providing an insert which is thermally conductive but electrically insulating has to be solved separately.

Ceramic (alumina) seals are attractive because of their good thermal conductivity, high dielectric strength, high compressive strength, and suitability as a substrate for deposited-film bridges, but their low temperature coefficient of expansion makes it difficult to achieve good seals between ceramic and body and between ceramic and connector pins; consistent difficulty appears to arise in attempts to produce good ceramic-to-metal brazes.

Because of the promising features of ceramic, a contract has been let for development and optimization of a ceramic seal design. This seal may prove to be a purely mechanical one (e.g., shrink fit) in which brazing or similar materials may be used for caulking or pressure-distribution rather than for bonds, or it may prove to be a "hybrid" seal in which glass either provides some of the strength or is used as a caulk. This contract is expected to be completed about September 1965.

A contract has also been let, and completed, for a literature search and tabulation relating to the temperature-degradation and other salient characteristics of suitable explosive materials.

Backup work on seals and preliminary work on deposited-film bridges is being done at JPL, as well as preliminary work on a nonproprietary match-head.

Prototype ceramic seals have already given encouraging results, but yield of squib threads with consequent extrusion of O-ring seals at 30,000 psi has served to illustrate the problems associated with attempting to seal pressures which approach the yield strength of many common alloys.

4. Sterilizable Primers

There is reason to believe that some spacecraft explosively actuated devices might better be percussion-initiated rather than electrically initiated, and the particular need here would be for sterilizable primers.

Frankford Arsenal reports that no primers (domestic or foreign) show more promise than two "400°F" primers (G-11 and G-16) developed for them by Remington. However, anomalous test results for primers tested at 300°F for JPL have been reported, and arrangements are consequently being made to test up to 1,000 of each type at JPL; until these tests are complete the pessimistic view must be taken that reliable sterilizable primers are not available.

5. Developmental Program

Once the seal design has been completed, the balance of the squib developmental program should be straightforward, with two exceptions: (1) the development of deposited-film bridges, and (2) anti-static-discharge features.

Although conventional bridgewires (typically nichrome of about 1.75 mil diameter) are reasonably satisfactory, the high ratio of perimeter-to-area and uniform thermal contact with a ceramic substrate make deposited-film bridges technically attractive. Past efforts to produce satisfactory deposited-film bridges have not been successful, probably because of separation between pins and film as a result of thermal stresses. Nevertheless, recent refinements in thin-film technology, as stimulated by the needs of electronic microcircuitry, now justify a further effort to produce a good deposited-film bridge for squibs. Apart from preliminary work being done at JPL, it is understood that at least one ordnance company has announced the production of a reliable unit, and assistance would also undoubtedly be available from those associated with the electronic industry.

Many approaches have been made to the problem of electrostatic discharge, including the provision of a simple shunt spark-gap at the connector end of the squib. If hazardous electrostatic charges could be built up in space, such a gap would have to be pressurized to be effective. One alternate to a gap is provision of low-resistance shunts, but here the problem is that of ensuring that the resistance of such shunts is not adversely affected by successive discharges. As yet there is no ideal solution to the electrostatic problem, but it is hoped that a method for producing a stable conducting surface on the ceramic can be developed.

References

1. Anderson, F. A., Strand, L. D., and Strehlow, R. A., *An Experimental Investigation of the Low-Pressure Combustion Limits of Some Solid Propellants*, Technical Memorandum No. 33-134, Jet Propulsion Laboratory, Pasadena, California, June 3, 1963 (Confidential).

2. Strand, L. D., "Low-Pressure Combustion Studies," *SPS 37-27*, Vol. IV, pp. 49–52, Jet Propulsion Laboratory, Pasadena, California, June 30, 1964.

XIII. Polymer Research

A. Equilibria Between Metallic Sodium and Aromatic Hydrocarbons

A. Rembaum, A. Eisenberg[1], and R. Haack

In the course of the investigation of the degradation kinetics of vinyl aromatic polyradicalanions (Ref. 1) a pronounced temperature effect was observed on the visible spectra of the poly(vinylbiphenyl) and poly(N-vinyl carbazole)-sodium complexes, as well as of the complexes of the corresponding ethyl-substituted aromatic hydrocarbons. The temperature effect could not be attributed to variation of volume changes due to thermal contraction or expansion, since the apparent extinction coefficient and also the free spin concentration as determined by the electron spin resonance technique (Ref. 1) varied by as much as a factor of six to seven.[2] Furthermore, the wavelengths of the absorption peaks remained unchanged, suggesting that no new complexes were being formed. These facts are therefore consistent with the postulated equilibrium (Ref. 2) between sodium and some aromatic hydrocarbons; the equilibrium was confirmed when a precipitation of metallic sodium was observed upon warming to room temperature a tetrahydrofuran solution of [N-ethyl carbazole]$^-$ or [4-ethyl biphenyl]$^-$-sodium radical anions prepared at $-70°C$, the precipitation being accompanied by a significant lightening of color. The phenomenon is completely reversible, with the sodium particles dissolving and the blue color darkening upon cooling.

Although equilibria between alkali metals and aromatic hydrocarbons have been considered previously in the study of polarographic redox potentials (Ref. 2), the influence of temperature on the extinction coefficients of charge transfer bands exhibited by hydrocarbon radical anions has, to our knowledge, never been studied, and the temperature-induced reversible precipitation of an alkali metal associated with an aromatic radical anion has not been observed.

A reversible color change was also noted by Weissman (Ref. 4) during the study of the ESR spectra of paracyclophane radical anions, but no precipitation of the

[1] Assistant professor in the Chemistry Department at the University of California at Los Angeles.

[2] The loss of paramagnetism with temperature was previously assumed (Ref. 3) to be a result of a disproportionation reaction and formation of dinegative ions. Our data indicate that this interpretation does not apply to the system described here.

alkali metal was reported, and apparently no quantitative investigations of the temperature effect were carried out.

In this communication the preliminary results of a quantitative study of the equilibria between sodium and some substituted aromatic hydrocarbons by spectroscopic techniques are summarized, and the possibility of acquiring new information regarding these equilibrium systems is indicated.

In a typical experiment, a solution of the aromatic species in tetrahydrofuran (10^{-3} molar) was kept on sodium at −70°C for a time sufficient to allow all aromatic nuclei to acquire an extra electron. It was then separated from the sodium and filtered while still at that low temperature, and sealed off from the sodium flask. The optical cell (optical path, 50 μ) was then filled and the absorbance measured. Throughout this procedure the sample was kept under vacuum.

The absorption spectrum of the N-ethyl carbazole recorded at three different temperatures in the wavelength range of 300 to 500 mμ is shown in Fig. 1. The charge transfer band at 390 mμ shows a remarkable change in temperature, while the bands at 350 mμ and below change only slightly.[3] The latter originates from electronic transitions which are not affected by the extra electron, in agreement with the interpretation given previously (Ref. 5). The variation of the absorbance at 390 mμ with temperature is shown in Fig. 2 and depicts more clearly the drastic change in absorbance which occurs within a temperature range of about 67°C.

The value of the equilibrium constant at various temperatures can be determined from Fig. 2 even without the knowledge of the molar extinction coefficient of the radical anion, assuming that Beer's law holds. Since a sodium titration has shown that every aromatic unit has acquired an electron at the lowest temperature, the low temperature absorbance plateau (A_0) minus the base line absorbance (A_B) is proportional to the total concentration of aromatic molecules, while A_t, the absorbance at any temperature, minus A_B, is proportional to the concentration of aromatic radical anions at that temperature. The equilibrium constant K thus becomes

$$K = \frac{A_t - A_B}{A_0 - A_t} \tag{1}$$

Plots of the values of this equilibrium constant as a function of temperature for three different species are shown in Fig. 3. Excellent linearity is observed over wide concentration ranges, confirming the ion-pair concept.

[3]The spectrum between 300 and 350 mμ (Fig. 1) is identical to the spectrum of unreacted N-ethyl carbazole in this wavelength range.

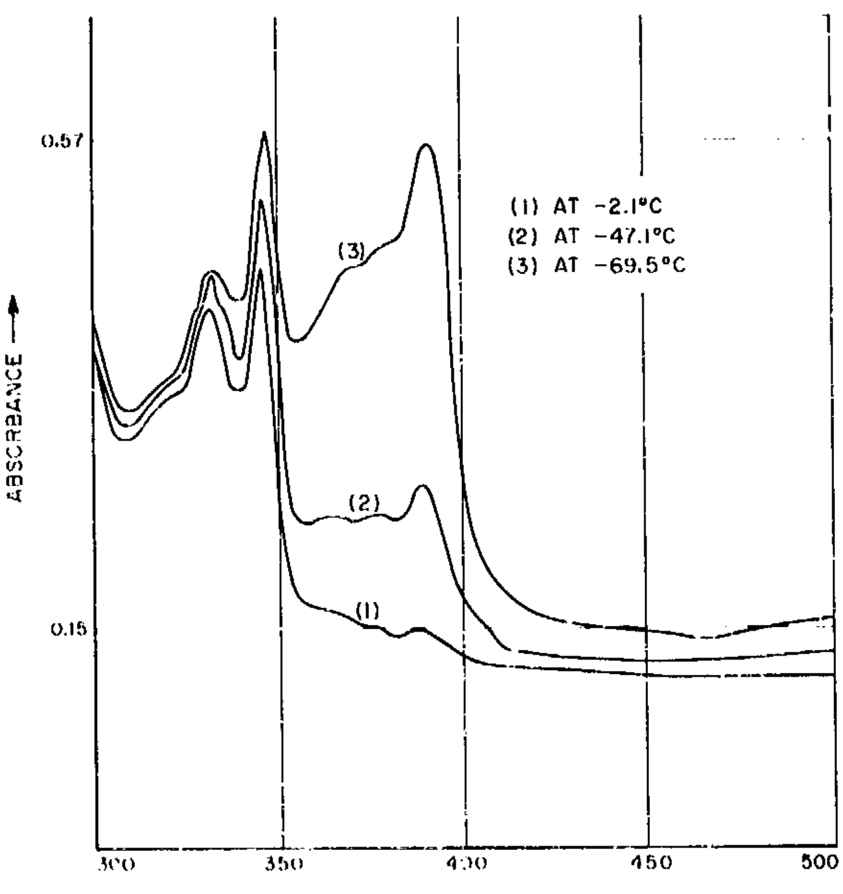


Fig. 1. Absorption spectrum of [N-ethyl carbazole]⁻ Na⁺ at three different temperatures

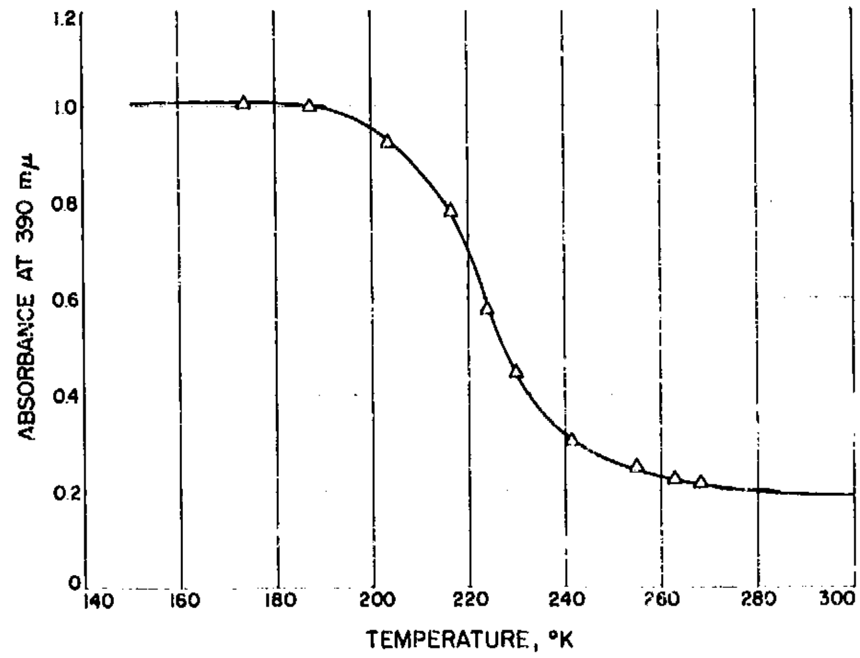


Fig. 2. Absorbance of N-ethyl carbazole radical anion versus temperature

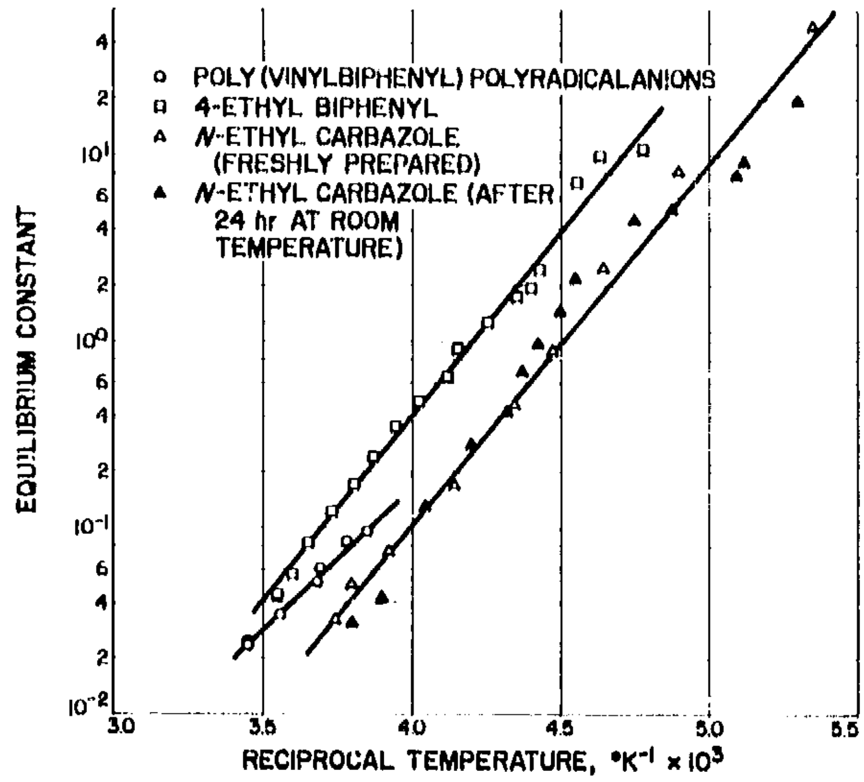


Fig. 3. Log of equilibrium constant versus reciprocal temperature

The calculated enthalpy changes ($\Delta H°$ values) are −9200, −9000, and −7000 cal/mole for 4-ethyl biphenyl, N-ethyl carbazole, and poly(4-vinylbiphenyl) radical anions, respectively.[4]

The benzene-sodium equilibrium has not yet been investigated by the spectrophotometric method described here. However, the well-known difficulty of forming the benzene radical anion at room temperature can now be explained. The above results with N-ethyl carbazole and 4-ethyl biphenyl indicate that the benzene radical anion probably exhibits a similar behavior to N-ethyl carbazole as a function of temperature, but its characteristic curve, which should resemble Fig. 2, is probably displaced to a low-temperature region. By contrast, the relatively small variations of the extinction coefficient in the range of −60 to 30°C of the spectrum of the naphthalene or biphenyl radical anion (Ref. 1) probably indicate that their characteristic curves lie above room temperature.

By the procedure outlined above, one can determine the equilibrium as a function of temperature for the reaction between alkali metals and various aromatic hydrocarbons in a range of solvents. The changes in enthalpy and entropy obtained from the equilibrium constants can in turn be correlated with the ionization potential of the alkali metal, the calculated electron affinity of the aromatic hydrocarbon, the dielectric constant of the solvent, etc. Finally, the technique offers a relatively simple method for preparing very finely divided sodium particles in a number of solvents, a phenomenon which may have applications in organic syntheses and in polymer chemistery.

[4]The calculated entropy change for the N-ethyl carbazole sodium equilibrium is unusually large (of the order of −40 e.u.) and this could be explained in terms of solvation effects (Ref. 1).

B. The Dependence of Resistivity on the Composition of the Poly(N-Vinyl Carbazole)-Tetracyanoethylene Charge Transf r Complex

J. W. Farrar and J. Moacanin

A previous study showed that the 1:1 poly(N-vinyl carbazole)-tetracyanoethylene (PVCA-TCNE) charge-transfer complex exhibits semiconductor properties, although the quantitative agreement between results obtained by different methods was unsatisfactory (*SPS 37-31*, Vol. IV). The purpose of this report is to present results of more quantitative measurements which have been carried out over the complete composition range of the complex. Modifications made on the Bridgman anvil apparatus for resistivity measurements (*SPS 36-14*, Vol. IV) led to marked improvements in the reproducibility of results.

The principal conclusions of this study are: (1) the enhanced conductivity and ohmic behavior for the complex were confirmed over a wide composition range; (2) the observed dependence of resistivity on composition is consistent with the postulate of a direct relationship between the number of charge carriers and the complex concentration; and (3), the Bridgman anvil technique can yield reproducible results on amorphous polymeric materials.

1. Experimental

The c mplex was prepared in the solid state by thoroughly blending with a mortar and pestle powdered PVCA and TCNE in the desired proportions. Although this method is somewhat more tedious than blending by precipitation from solution, it has the advantage of obviating the possibility of contamination by traces of solvent or solvent impurities.

Resistivity measurements were carried out on compressed powders of the complex by means of a previously described Bridgman anvil apparatus (*SPS 37-31*, Vol. IV and Ref. 6). The following modifications were made in order to decrease noise and stray field pickup to a level below the sensitivity of the Keithley 600A meter. The standard wiring previously used was replaced with an RG/213 coaxial cable and VHF connectors. The box housing the voltage supply dry-cell batteries was lined with Teflon sheets. With these modifications stray pickup was virtually eliminated and the shunting leakage resistance was greater than 10^{13} Ω. The modified setup was checked by measuring the resistance of a Victoreen standard resistor (10^{11} Ω ± 10%) connected between the two anvil brackets. The determined *I-V* characteristic is given in Fig. 4. Inasmuch as the error for the meter is about ±4% of full scale, and allowing for a drop of about 1% per year for the standard resistor in the meter (Ref. 2), the results fall within this high side of the cumulative error limits. Thus it is safe to estimate the over-all accuracy of the resistance measurements in the range > 10^{11} Ω to be about ±25%.

2. Results and Discussion

Previously it was shown that the 1:1 PVCA-TCNE complex exhibits ohmic behavior up to fields of about 8000 v/cm (*SPS 37-31*, Vol. IV). The new results show that this behavior holds for the complete range of complex composition and for all the pressures at which measurements were made. The *I-V* data for 14 and 18 mole % TCNE are shown in Figs. 5 and 6, respectively. The resistivity data for the various pressures are summarized in Fig. 7. For convenience of presentation, results are given for only three pressures, although data were obtained for four additional intermediate pressures. Although the pressure dependence follows approximately the linear relation between log σ and $P^{1/2}$ as proposed by Pohl (Ref. 6), the results in this respect are inconclusive.

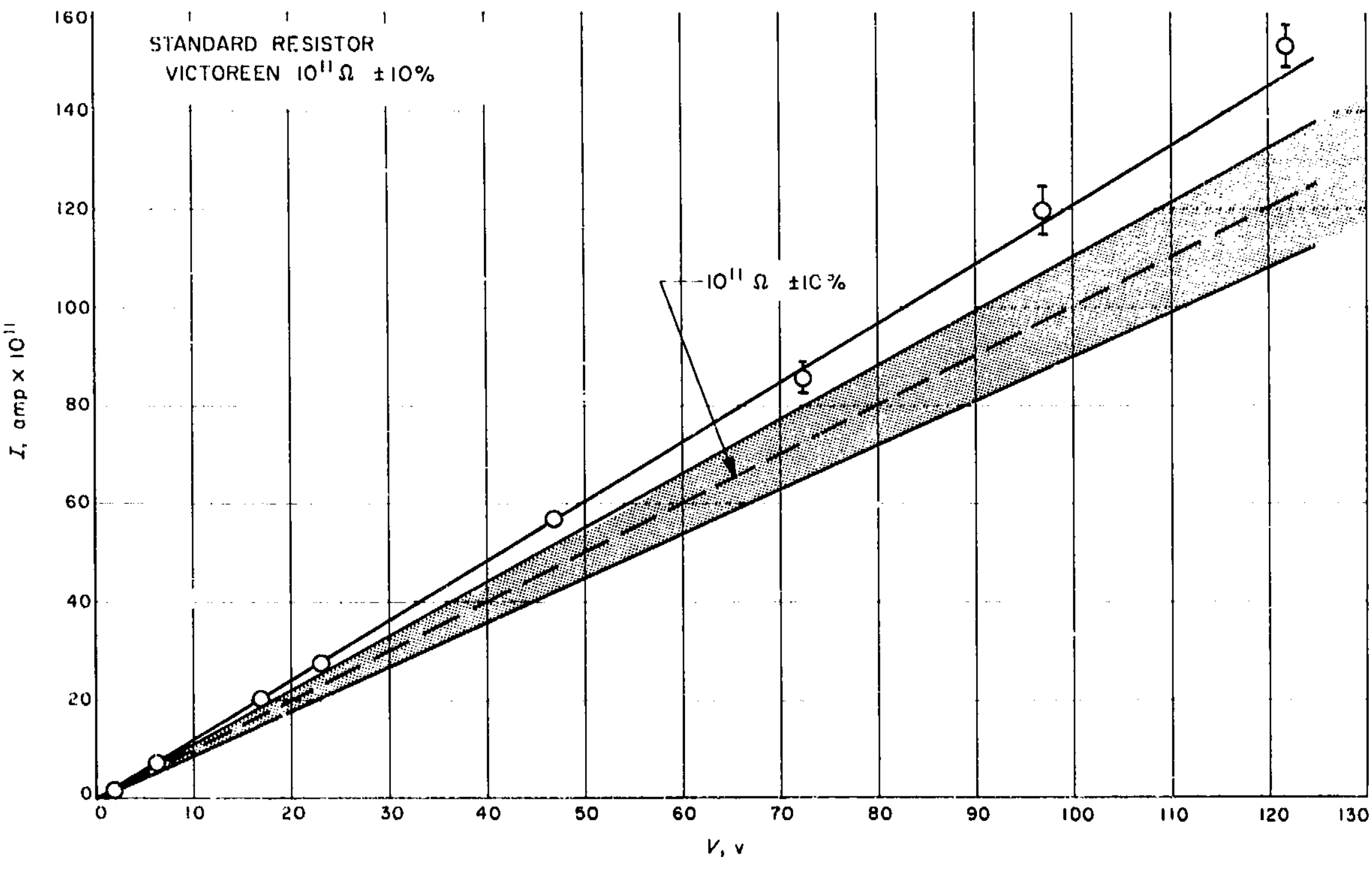


Fig. 4. Check of the performance of the Bridgman anvil apparatus with a standard resistor

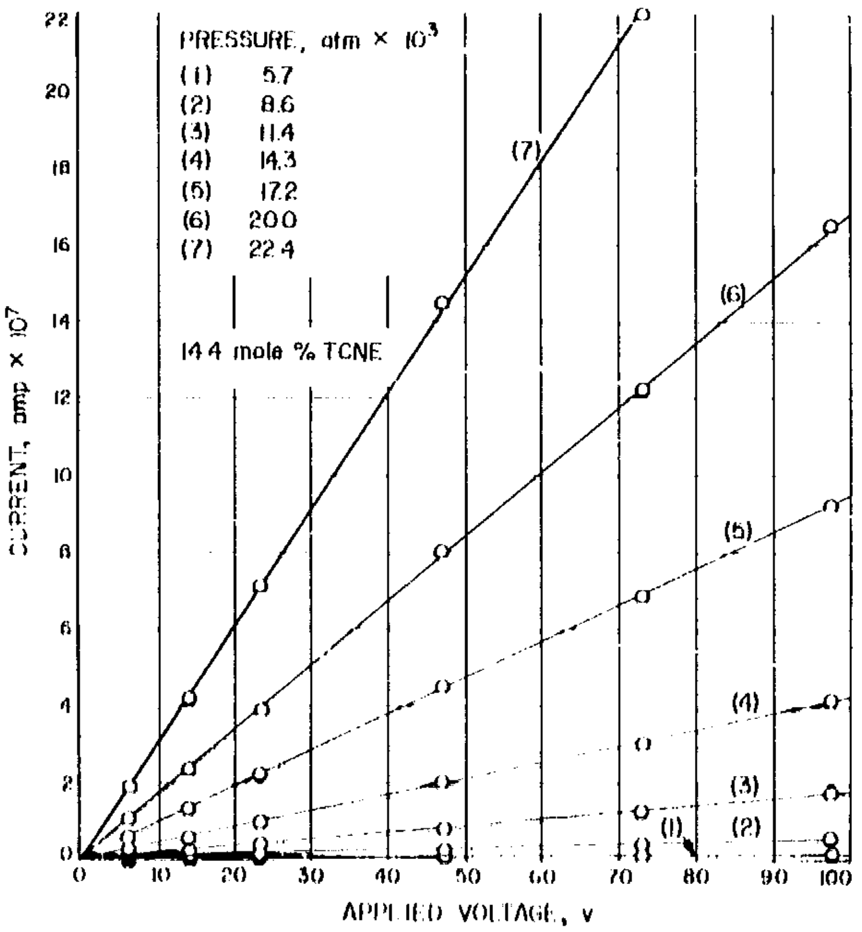


Fig. 5. Current-voltage characteristics for the complex containing 14.5 mole % TCNE

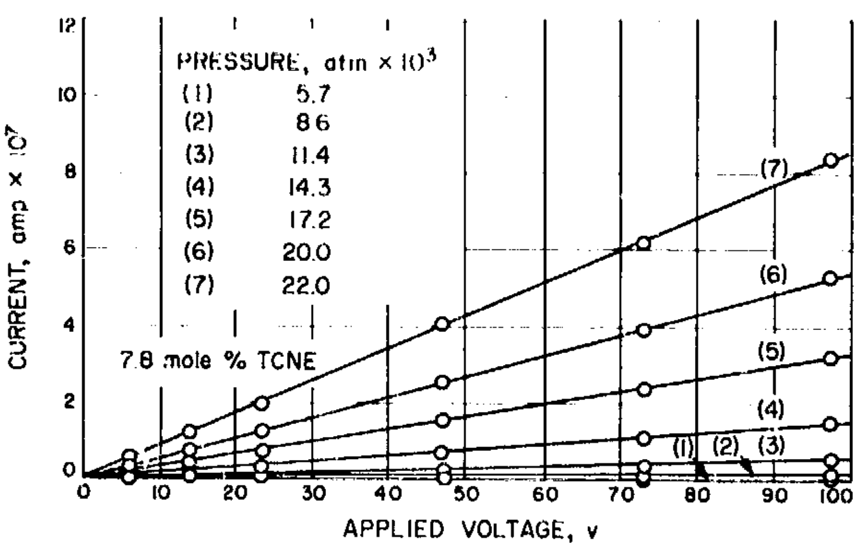


Fig. 6. Current-voltage characteristics for the complex containing 18 mole % TCNE

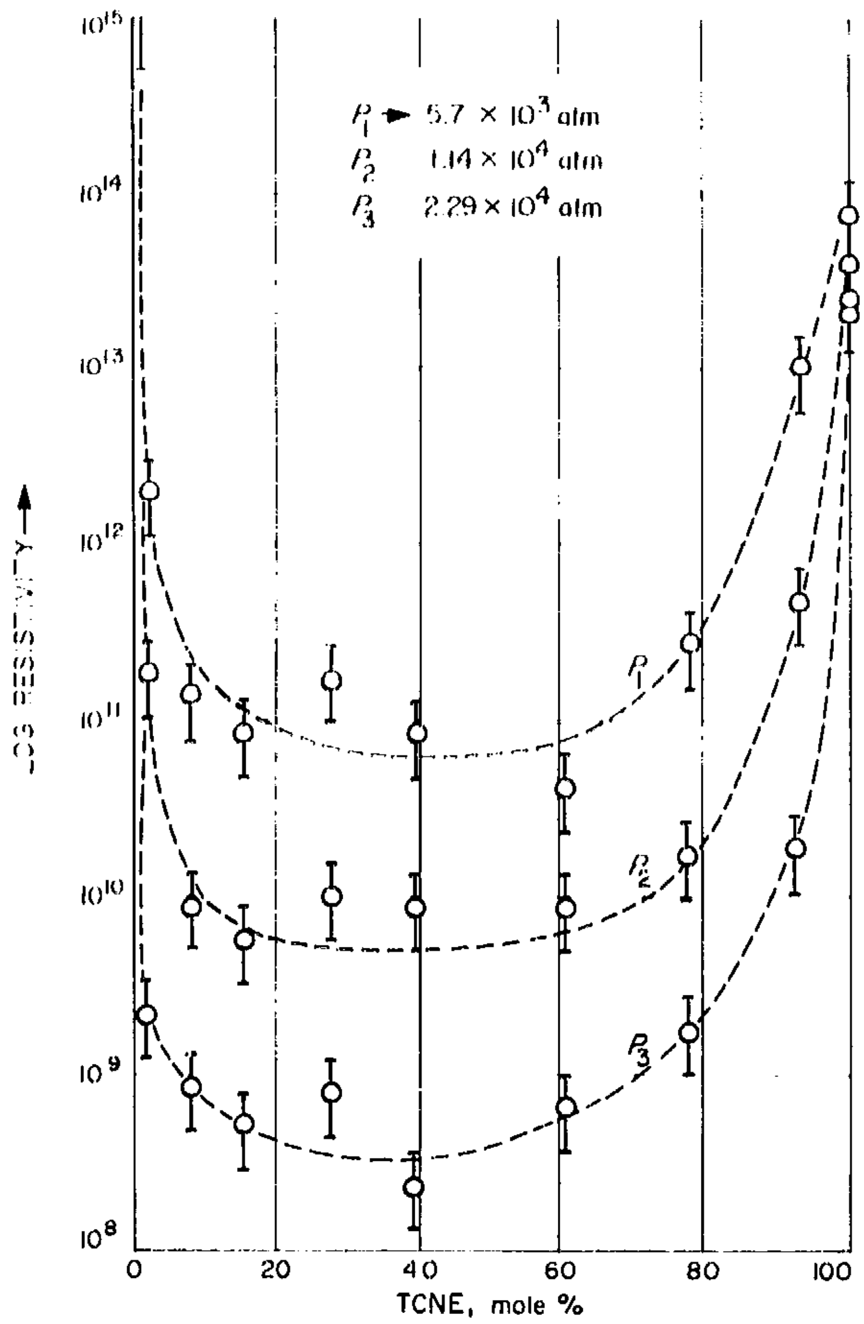


Fig. 7. Resistivity versus composition of PVCA–TCNE at various pressures

The resistivity of PVCA was above the upper range of the electrometer, i.e., $> 10^{14}$ Ω-cm, a result consistent with the reported value of 8×10^{15} at atmosphere pressure (Ref. 8). The value for TCNE is in reasonable agreement with the value of approximately 10^{13} Ω-cm reported by Slough (Ref. 9). The agreement with the literature results for the pure components also suggests that, for the experimental arrangement, used surface conductivity cannot be a serious factor because it would affect most strongly the high resistivity results.

The dramatic decrease in resistivity with addition of only 1.5 mole % TCNE is to be noted. Addition of 10% further decreases resistivity by a factor of about 15, but addition beyond this point does not seem to result in additional marked changes. Above 80 mole % TCNE the resistivity rapidly increases. This result is consistent with the concept that the number of charge carriers is proportional to the complex concentration. The relatively minor changes observed in the region of intermediate composition may merely reflect comparable changes in complex concentration, or by inference, in the number of charge carriers. Recent studies have shown that in solution an equilibrium, subject to the usual mass action law, exists between TCNE, the carbazole moieties, and the complex. This work is now being extended to the solid state.

In conclusion, it may be worthwhile to comment on the good reproducibility of the results using the Bridgman anvil technique. Most of the criticism of this method is applicable to measurements on polycrystalline materials where intergranular contacts have a significant effect. We have experienced similar difficulties with such materials. However, powders of amorphous materials such as PVCA will flow under pressure to form a homogeneous film.

C. Mechanical Behavior of Styrene-Butadiene Rubber Filled with Glass Beads

R. F. Fedors and R. F. Landel

1. Introduction

In analogy with the behavior of unfilled gum elastomers, the uniaxial stress-strain properties of filled rubbers, including the values of the stress and strain-at-break, can be expected to depend on many factors such as the rate and temperature of testing. It is also true that the majority of work, both experimental and theoretical, has been expended on the study of the gum elastomers, since the response of these systems seems intuitively simpler to investigate and interpret than that of the filled systems.

Studies of the response of the unfilled systems have demonstrated the utility of the failure envelope concept (Ref. 10). In addition, other studies (Ref. 11) have also emphasized the very important influence of parameters such as the concentration of effective network chains ν_e, and the number of statistical units per network chain n. Briefly, the failure envelope is generated for a given gum vulcanizate when the values of $\sigma_b T_0/T$ versus ϵ_b are plotted in log-log coordinates; σ_b and ϵ_b are the stress and strain-at-break respectively, T_0 is an arbitrary reference temperature and T, the test temperature. Of extreme importance is the fact that the failure envelope is rate and temperature independent.

In addition, studies of the effect of varying ν_e on the location and shape of the failure envelope have also been made (Ref. 11). It has been demonstrated that superposition of the high-temperature segments of individual failure envelopes representing gum elastomers of widely differing chemical structure and differing ν_e values can be effected by reducing the quantity $\sigma_b T_0/T$ to unit ν_e, i.e., by plotting $\sigma_b T_0/\nu_e T$ versus ϵ_b. Since different elastomers with widely varying T_g values were being compared, T_0 was taken to be $T_g + 100$. Fig. 8 shows data for some 15 systems superposed in this manner (Ref. 11). The figure is more significant than merely a demonstration of a method which permits a correlation of ultimate property data for elastomers of widely differing chemical structure. Of far greater importance is the fact that it demonstrates that the ultimate properties of gum elastomers depend primarily on the network-like structure of the rubber and not on the detailed chemical constitution along the chain, except as this influences n. Thus the superposed data in Fig. 8 delineate a characteristic failure curve which gives the breaking strength per chain in the network. To find the actual breaking properties over a large portion of the time-temperature scale, then, it is only necessary to know the number of chains per unit volume of rubber, i.e., the density of crosslink sites.

Furthermore, it has also been shown that the value of n is an important parameter, since the maximum value of the extension ratio-at-break $(\lambda_b)_{max}$, $(\lambda = \epsilon + 1)$ varies as $n^{1/2}$ (Ref. 11). Since $n \propto \nu_e^{-1}$, $(\lambda_b)_{max}$ varies directly as $\nu_e^{-1/2}$. Therefore a knowledge of ν_e is sufficient to ascertain the permissible range of values which both σ_b and ϵ_b can assume for a gum rubber.

In view of this success with the gum system, it seems worthwhile to attempt an application and extension of these methods to filled rubber composites. As might be expected, the presence of a filler in relatively large concentrations can considerably complicate the acquisition and interpretation of data. For example, in addition to the obvious variable of filler concentration, it is conceivable that other factors, such as the extent of filler-polymer adhesion, the filler size and shape, and the distributions of size and shape, could influence the behavior of the composite. It is apparent that these factors may produce effects which are not independent, and hence interpretation of the data would be extremely difficult if all factors were allowed to exert their influence simultaneously. Hence, the use of a model system where arbitrarily chosen variables can be studied while others are held constant is implied as a reasonable approach in a study of filled systems (Ref. 12).

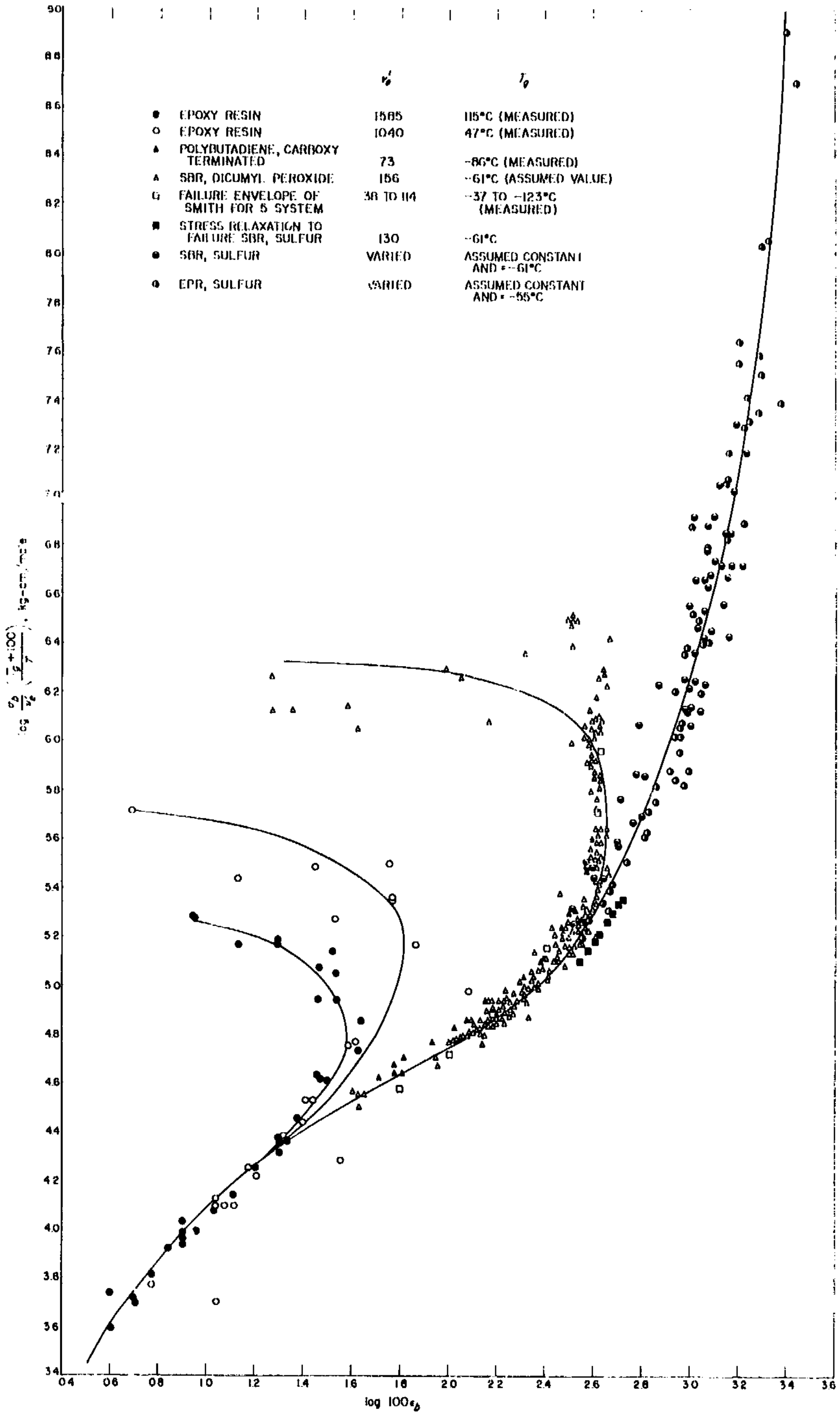


Fig. 8. Reduced failure envelopes for several amorphous gum elastomers

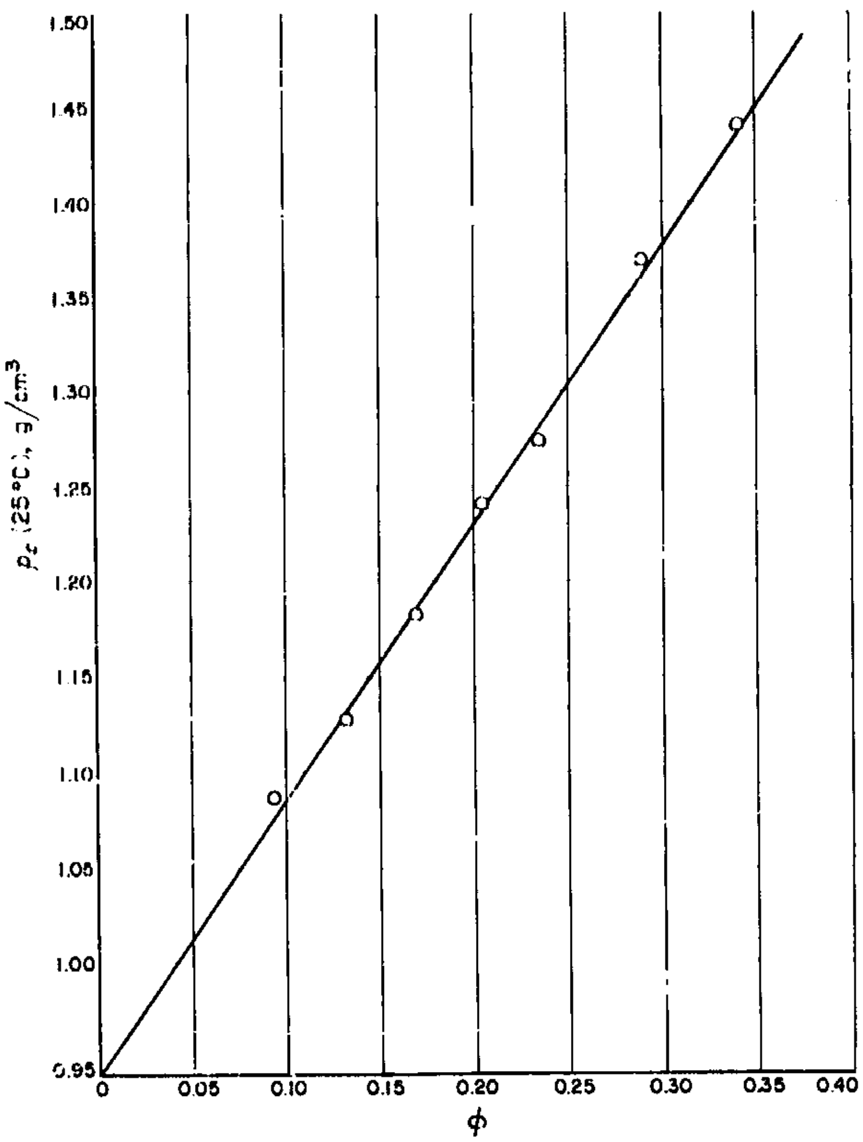


Fig. 9. Dependence of the density of the composite ρ_c on calculated volume fraction of filler ϕ

The model system chosen for preliminary study was a styrene-butadiene rubber (SBR) filled with varying amounts of glass beads and vulcanized by means of dicumyl peroxide. Glass beads were chosen as filler because these are commercially available in various sizes; in addition, it was hoped that the filler-polymer adhesion would be small if not completely absent.

It was also hoped that the glass would be inert and hence have no effect on the vulcanization reaction. Since the ratio of peroxide to rubber was held fixed, ν_e values would be constant and independent of the volume fraction of glass ϕ. This summary will describe the preparation and characterization of the samples of filled material, while a subsequent summary will describe the uniaxial stress-strain behavior.

2. Experimental

Styrene-butadiene rubber (Shell 1500) was used throughout this study as the base elastomer. The curing agent was dicumyl peroxide (Hercules Di Cup R). Glass beads (Superlite) were obtained from Minnesota Mining and Manufacturing Company. The reported number-average diameter of the essentially spherical filler particles was 29 μ. Direct measurement showed that the filler-size distribution could be described by the normal distribution function with a most probable value of the diameter equal to 34 μ. The compositions of the samples, by weight, are as shown in Table 1.

The compositions were mixed on a 2-roll rubber mill using normal mixing techniques. Care was taken to maintain the stocks at a reasonable temperature during incorporation of the glass filler. The temperature tended to increase during mixing, especially with the compositions having high filler loadings. Vulcanization was accomplished in an ASTM mold under pressure at 325°F for 60 min.

Equilibrium swelling in benzene at 25°C was utilized to estimate the ν_e values for cured samples. Small specimens, dried to constant weight, were allowed to swell in benzene in the dark until no further swelling occurred. This required about one week during which the solvent was changed at least four times. After the specimens had reached maximum swelling as determined by weight, they were dried to constant weight. The difference in the initial dry weight before swelling and the final dry weight after swelling was taken as the weight of soluble rubber.

Densities of the composites were determined at 25°C by means of pycnometers, using water as the confining fluid.

Table 1. Composition of filled SBR samples

Designation	CA	3A	4A	5A	6A	7A	8A	9A	10A
SBR (Shell 1500)	100	100	100	100	100	100	100	100	100
Dicumyl peroxide	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57
Glass	0	26.2	39.3	52.4	65.4	78.8	92.2	105	132

3. Results and Discussion

a. Homogeneity of the composites. Experience shows that remarkable agreement is generally observed between the measured volume and a value calculated on the assumption of simple additivity of the volume of each component for systems of the type studied here, provided that the composites are reasonably free of voids. Based on this assumption, the density of the composite ρ_c, is related to the density of the rubber ρ_r, and the density of the glass filler ρ_g by the following relationship

$$\rho_c = \rho_r + \phi\,(\rho_g - \rho_r) \tag{1}$$

where ϕ is the calculated value of the volume fraction of glass in the composite and is obtained from

$$\phi = \frac{w_g/\rho_g}{w_g/\rho_g + w_r/\rho_r} \tag{2}$$

where w_g and w_r are the weights of the glass and rubber, respectively, in the composite. In Fig. 9 the density of the composites is shown as a function of ϕ. The linear relationship predicted by Eq. (1) is evidently valid. According to Eq. (1) the slope of the line in the figure corresponds to $(\rho_g - \rho_r)$. The predicted value of this quantity obtained from the figure is 1.461. The measured densities of the glass and the unfilled rubber yield a value of 1.477 for $(\rho_g - \rho_r)$. The close agreement between Eq. (1) and the data shown in Fig. 9 establishes that the undeformed composites are reasonably free of measurable voids.

b. Crosslink concentration. An attempt was made to keep the crosslink concentration constant and independent of ϕ by maintaining the ratio of rubber to peroxide constant. In view of previous studies (Ref. 11) which indicated the significant effect which ν_e has on both the stress-strain properties and the values of σ_b and ϵ_b, it is important that ν_e be known for each composite.

For unfilled rubbers, ν_e can be easily calculated from values of the volume fraction of rubber in a swollen gel v_2, at equilibrium swelling. These required v_2 numbers are obtained from

$$v_2 = \frac{w_r/\rho_r}{w_r/\rho_r + w_l/\rho_l} \tag{3}$$

where w_l and ρ_l are the weight of imbibed liquid at equilibrium and the density of the liquid, respectively. The ν_e value can then be calculated from the Flory-Rehner (Refs. 13, 14) equation

$$\nu_e = -\frac{g}{v_0}\left[\frac{\ln(1 - v_2) + v_2 + \chi_1 v_2^2}{v_2^{1/3} g^{2/3} - v_2/2}\right] \tag{4}$$

where v_0 is the molar volume of solvent, g is the gel fraction and χ_1 is the solvent-polymer interaction parameter.

However, in the case of filled systems, the situation is more complicated because the connecting link between v_2 and ν_e, Eq. (4), is strictly applicable to gum rubbers only. In order to be able to use Eq. (4) for filled systems, it is necessary that the v_2 values which are calculated for filled composites be independent of the presence of the filler and thus correspond to the values which would be obtained for the analogous gum vulcanizates.

Kraus (Ref. 15) has treated this problem of the swelling of filled composites theoretically and, based on certain reasonable simplifying assumptions, derived the following equations

$$\left[(1 - \phi)/\phi\right]\left[1 - (v_2/v_2^*)\right] = v_2 + 1 - 3c\,(1 - v_2^{1/3}) \tag{5}$$

where v_2 is the true volume fraction of rubber (i.e., the volume fraction which would have been observed if no filler were present), v_2^* is the apparent volume fraction of rubber and is given by

$$v_2^* = \frac{(w_c/\rho_c)\,(1 - \phi)}{(w_c/\rho_c)\,(1 - \phi) + w_l/\rho_l} \tag{6}$$

Here w_c is the weight of the composite as measured after swelling and drying. The quantity c is a parameter, independent of both ϕ and v_2, which is a measure of the ability of the filler particle to restrict swelling of the rubber matrix in the neighborhood of the filler surface (or of a nonuniform distribution of crosslinks in the neighborhood of the filler surface). If there is no adhesion between the filler and binder, then swelling of the rubber matrix in the composite will be normal, except that each filler particle will be surrounded by a cavity containing the swelling liquid. When this is true, $c = 0$. This case would correspond roughly to the swelling of a closed-cell foam; i.e., the liquid-filled cells in the swollen foam would be analogous to the cavities containing filler and pure liquid in the swollen composite case.

When $c = 0$, Eq. (5) becomes

$$1/v_2^* = (1/v_2) + \left[(1 - v_2)/v_2\right]\left[\phi/(1 - \phi)\right] \tag{7}$$

As the degree of adhesion between the binder and filler increases (or as the distribution of crosslinks becomes more nonuniform in the neighborhood of the filler) the value of c will increase. Kraus (Ref. 15) has tabulated c

values for several types of carbon black filler; c varied from a low of 0.77 for a nonreinforcing graphitized carbon black in butyl rubber to a high of 1.35 for a highly reinforcing acetylene carbon black in SBR.

For the present SBR-glass bead system, the calculated and derived physical property data are shown in Table 2. Considering the values of the gel fraction g, an invariable increase of g with increase of ϕ strongly suggests that ν_e also increases with ϕ, despite the fact that the ratio of peroxide to rubber was held constant. This would imply that the filler increased the efficiency of the peroxide crosslinking reactions. This effect is supported by a report which shows that basic fillers (glass beads in the present case) generally cause an increase in the efficiency of dicumyl peroxide crosslinking (Ref. 16).

If the variation in ν_e with ϕ is real, then it is not possible to use Eq. (5) directly in order to obtain the true volume fraction v_2. However, by making an assumption about the value of c, it is possible to obtain values of v_2, which may then be combined with Eq. (5) to calculate a value of c; a comparison of the assumed value and the calculated value will indicate the degree of validity of the original assumption. For SBR-glass beads, it seems reasonable to suppose $c = 0$ as a first approximation.

For this case, Bills and Salcedo (Ref. 17) have pointed out that the true volume fraction v_2 can be calculated from the measured swelling behavior of composites by means of

$$v_2 = \frac{w_r/\rho_r}{w_r/\rho_r + w_l/\rho_l} \qquad (8)$$

Table 2. Physical properties of SBR-glass bead composites

Sample designation	ϕ	ρ_r (25°C), g/cm^3	gel fraction	v_2	v_2*	c	$\nu_e \times 10^6$ moles/cm^3
CA	0	0.9430	0.887	0.130	—	—	102
3A	0.0922	1.086	0.926	0.146	0.135	0.005	111
4A	0.132	1.128	0.934	0.145	0.129	0.004	108
5A	0.169	1.182	0.942	0.148	0.126	0.005	115
6A	0.204	1.240	0.947	0.155	0.128	0.003	124
7A	0.234	1.273	0.951	0.151	0.120	0.001	121
8A	0.270	1.329	0.954	0.157	0.120	0.004	131
9A	0.289	1.369	0.958	0.160	0.120	0.002	133
10A	0.339	1.439	0.960	0.163	0.114	0.005	142
Glass beads	—	2.42	—	—	—	—	—

Table 3. Physical properties of silicone rubber-glass bead composites

Sample designation	ϕ	ρ_r (25°C)	gel fraction	v_2	v_2*	c
29/μ-D beads						
1	0	1.038	0.940	0.388	—	—
3	0.05	1.110	0.953	0.435	0.422	−0.005
4	0.10	1.181	0.952	0.405	0.380	−0.030
5	0.20	1.310	0.956	0.431	0.377	−0.005
6	0.30	1.447	0.972	0.516	0.427	+0.010
7	0.50	1.719	0.973	0.563	0.392	+0.004
50/μ-D beads						
2	0	1.038	0.954	0.439	—	—
8	0.05	1.110	0.945	0.385	0.373	+0.010
9	0.10	1.181	0.948	0.392	0.367	0.020
10	0.20	1.310	0.958	0.443	0.387	−0.020
11	0.30	1.447	0.971	0.507	0.419	+0.005
12	0.40	1.719	0.980	0.614	0.443	0.000

where w_l represents the total weight of imbibed liquid, including the liquid present in the cavities surrounding each filler particle. Calculations of v_2 based on Eq. (8) and of the corresponding quantity v_2^* based on Eq. (6) are listed in Table 2. The c values obtained from Eq. (5) are also listed, and it is evident that within experimental error, $c = 0$, which justifies the initial supposition concerning the value of c. Hence, the v_2 values from Eq. (8) can be used directly with Eq. (4) ($\chi_1 = 0.34$) to calculate ν_e, which values are also shown in Table 2. Thus, it has been shown that the general increase of ν_e with ϕ is a real effect due presumably to the participation of the glass filler in the peroxide vulcanization reaction such that the efficiency of crosslinking reactions is increased. The results obtained also indicate that the increased number of crosslinks are not confined to the neighborhood of the particle surface but rather are distributed uniformly throughout the rubber matrix.

The assumption that $c = 0$ for glass beads is supported by data obtained by B. G. Moser[1] from equilibrium swelling studies in benzene on composites composed of a silicone binder with glass bead filler. In this case two sizes of glass were used, 29- and 50-μ number-average diameters. The calculated and derived data are shown in Table 3. The values of v_2 and v_2^* were again obtained from Eqs. (8) and (6), respectively. For both sizes of glass, it is evident that the value of c can be taken to be zero within experimental error.

[1] Private communication.

References

1. Rembaum, A., et al., "Electron Transfer to Vinyl Aromatic Polymers" (to be published).
2. Hoijtink, G. J., de Boer, E., Van der Meij, P. H., and Weijland, W. P., *Recueil des Travaux Chimiques*, Vol. 74, p. 277, 1955; Ibid, Vol. 75, p. 487, 1956.
3. deBoer, E., *Advances in Organometallic Chemistry*, Vol. 2, Academic Press, New York, N. Y., 1964.
4. Weissman, S. I., *Journal of the American Chemical Society*, Vol. 80, p. 6462, 1958.
5. Paul, D. E., Lipkin, D., and Weissman, S. I., *Journal of the American Chemical Society*, Vol. 78, p. 116, 1956.
6. Pohl, H. A., Rembaum, A., and Henry, A., *Journal of the American Chemical Society*, Vol. 84, p. 2699, 1962.
7. Instruction manual for Model 600A Multi-Purpose Electrometer, Keithley Instruments, Inc., Cleveland, Ohio.
8. Busse, W. F., Lambert, J. M., McKinley, C., and Davidson, H. R., *Industrial and Engineering Chemistry*, pp. 2271–2275, December 1948.
9. Slough, W., *Transactions of the Faraday Society*, Vol. 58, p. 2360, 1962.
10. Smith, T. L., *Journal of Polymer Science*, Part A, Vol. 1, p. 3597, 1963.
11. Landel, R. F., and Fedors, R. F., *Transactions of the Society of Rheology*, Vol. 8, (in press); Landel, R. F., and Fedors, R. F., Chapter IIIB, *Fracture Processes in Polymeric Solids*, B. Rosen, ed., J. Wiley and Sons, New York, 1964.
12. Landel, R. F., *Transactions of the Society of Rheology*, Vol. 2, p. 53, 1958.

References (Cont'd)

13. Flory, P. J., and Rehner, J., *Journal of Chemical Physics*, Vol. 11, p. 521, 1943.

14. Flory, P. J., *Principles of Polymer Chemistry*, Chapter XII, Cornell University Press, 1953.

15. Kraus, G., *Journal of Applied Polymer Science*, Vol. 7, p. 861, 1963.

16. Hercules Powder Company, Technical Report 102, March 1962, Wilmington, Del.

17. Bills, K., and Salcedo, F. S., *Journal of Applied Physics*, Vol. 32, p. 2364, 1961.

XIV. Research and Advanced Concepts

A. Spectroscopic Measurement of a Detached Shock Wave in an Arc-Heated Plasma Jet

C. J. Chen

The study of the structure of shock waves in a plasma has received much attention in the last few years. Investigators, as Denesse and Rocard (Ref. 1), Sen (Ref. 2) and Marshall (Ref. 3) focused their attention on fully ionized equilibrium gases. Later, Jukes (Ref. 4), Kolbel (Ref. 5), Tidman (Ref. 6), Shafranov (Ref. 7), Zel'Dovich (Ref. 8), Imshennik (Ref. 9), and Prokofev (Ref. 10) removed the restriction of thermal equilibrium. Most of the investigators used Navier–Stokes-type equations for different species together with the charge neutrality assumption to determine the distribution of the parameters across the shock. The general feature of their results indicates that the electron temperature changes gradually across the shock due to the fact that the electrons have a very high thermal conductivity, while charge number density and ion temperature undergo a Rankin-Hugoniot type of "jump."

The problem of shock properties in partially ionized gases is treated by Grewal and Talbot (Ref. 11). The authors solve this problem first by considering the neutrals and ions as one species. The Mott-Smith model is used to obtain the distribution of the flow parameters across the shock by assuming that the degree of ionization is small so that the energy perturbation of ion-neutral species caused by the electrons can be neglected. The mechanism affecting the electron density is a compression and ionization. Owing to the extremely high thermal conductivity of the electrons, the electron temperature profile is broadened ahead of the shock.

Preliminary results of the shock processes in a partially ionized plasma obtained by using spectroscopic techniques are presented in this note. Fig. 1 shows the schematic diagram of the apparatus. The arc jet equipment consists of a commercial 45-kw arc facility and a 2.8 Mach nozzle of 1-in. exit diameter. A Jarrell-Ash 500-mm spectrometer is used to monitor the radiation of the plasma in the vicinity of the standoff shock in front of a hemisphere immersed in the plasma jet. To achieve

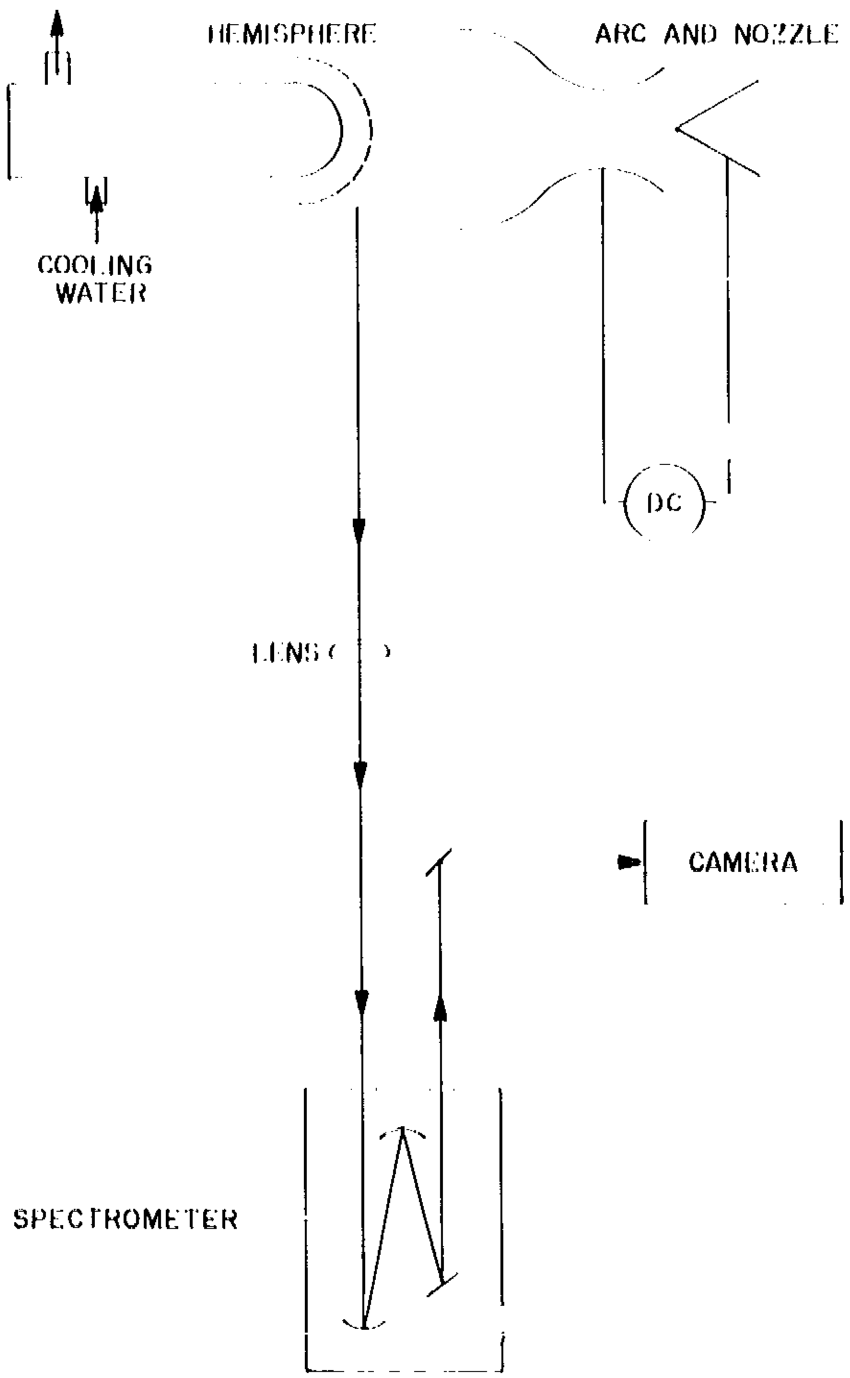


Fig. 1 Apparatus arrangement for measurements

high spatial resolution (~0.2 mm) an optical system of large f-number (~30) is used. A trace (~0.1% by volume) of hydrogen gas is introduced into the argon stream. The H_β line and the 4702.13A argon line are photographed both with the slit of spectrometer along the centerline of jet in the vicinity of the shock, and with the slit perpendicular to the centerline of the jet at various positions along the jet. Abel's integral equation is used to evaluate the centerline intensity at various positions along the axis of the jet. The H_β profile at the centerline of the jet is also obtained by the same processes. Griem's (Ref. 12) formulation is utilized to determine the electron density. The distribution of relative line intensity along the centerline of the jet is used to obtain the normalized electron temperature distribution across the shock.

The intensity of the argon line normalized to free-stream intensity can be written as

$$\frac{I}{I_0} = \frac{n}{n_0} \exp\left[\frac{E}{k}\left(\frac{1}{T_0} - \frac{1}{T}\right)\right]$$

Solving for T/T_0 we have

$$\frac{T}{T_0} = \left[\frac{kT_0}{E} \ln \frac{n}{n_0} \frac{I_0}{I} + 1\right]^{-1}$$

where I is the intensity of the line at any position, T the electron temperature, and E the upper energy level of the line. The quantities with zero subscripts refer to the free-stream conditions. T/T_0 is obtained by using the measured values of $(n/n_0)/(I_0/I)$.

The electron density distributions are shown in Fig. 2. The density jumps to about 7.5 times the free-stream condition within 1 mm spatial spread. Owing to the limitation of the spatial resolution the details of the distribution of plasma properties in the shock cannot be studied. The causes of this density jump are thought to be divided into two categories: the shock compression and further ionization due to the temperature rising behind the shock. The density behind the shock due to compression can be obtained from the Mach number of

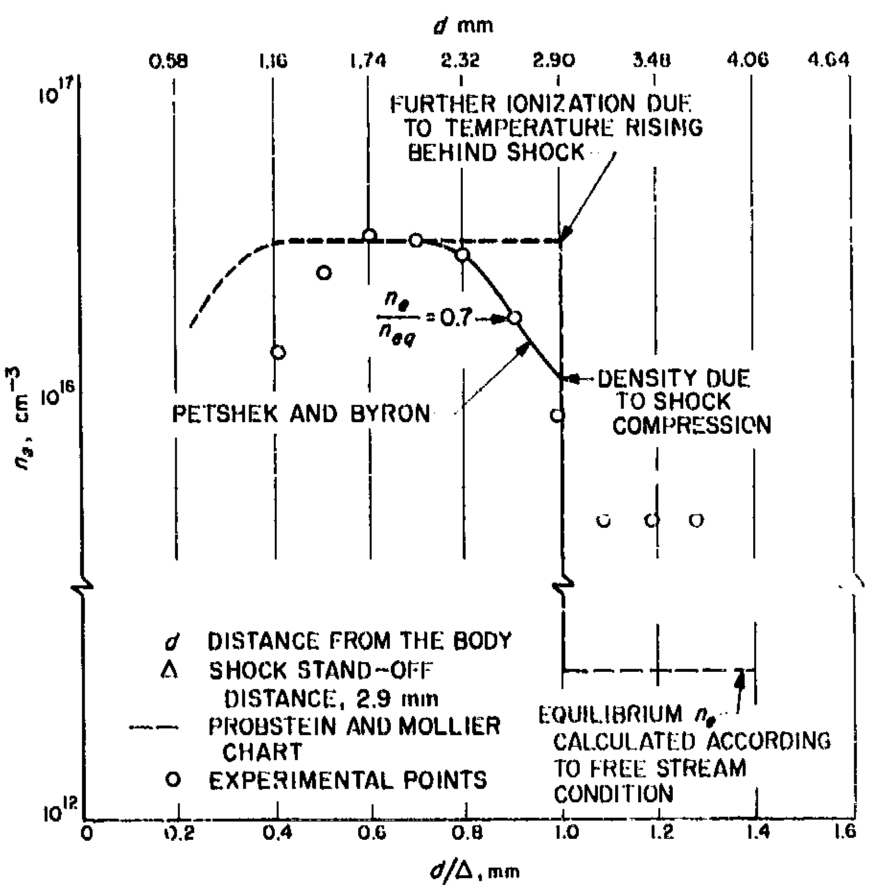


Fig. 2. Electron density distribution across shock

the shock. The mechanism of further ionization behind the shock as studied by Petschek and Byron (Ref. 13) and Jahn and Howard (Ref. 14) is used to interpret the data. The mechanism consists of the following: (1) The rate of ionization of the argon atom from the first excited state is much higher than the rate of excitation from the ground state to the first excited state. This rate of excitation is the limiting rate for the ionization process. (2) The electrons loose energy by inelastic collisions and gain energy by elastic collisions through electron-ion and electron-atom interactions. By equating the rate of energy transfer of the interactions they are able to obtain the rate of ionization depicted graphically as shown in Fig. 3. The data obtained in this experiment are fitted in this figure as indicated by the circles. The symbol τ in Fig. 3 refers to the time required for the gas to reach 70% of the equilibrium degree of ionization, n_e and n_{eq} are electron density and electron equilibrium density, respectively.

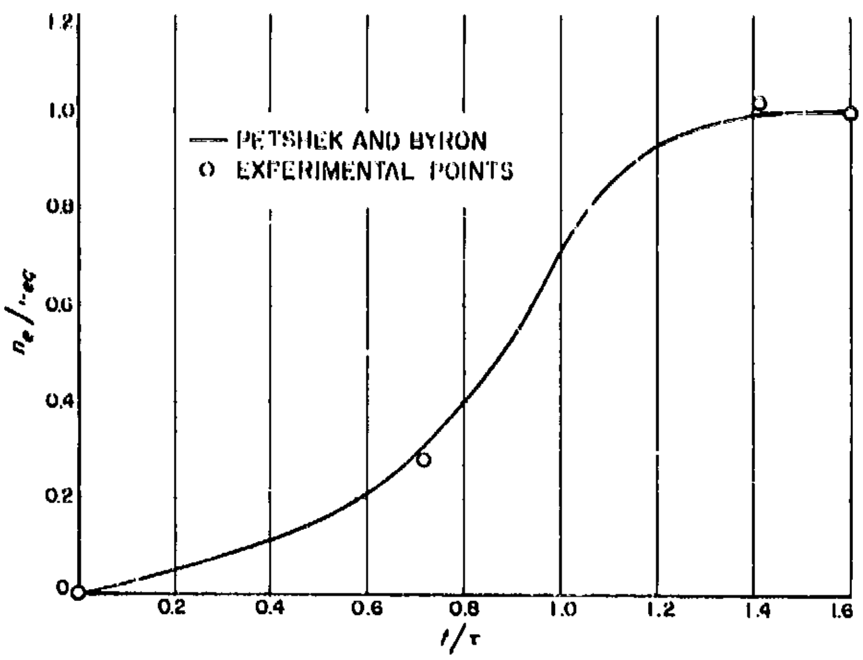


Fig. 3. Ionization time of gas behind shock

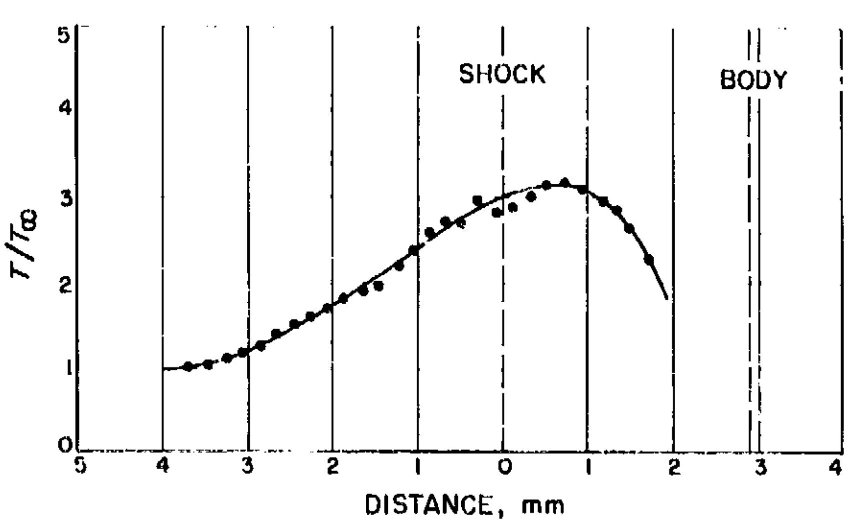


Fig. 4. Electron temperature distribution across shock

The electron temperature distribution obtained is shown in Fig. 4. There is no electron temperature jump across the shock. The distribution of the electron temperature is broadened to about 2000 free-stream mean free paths, if Spitzer's mean-free-path formulation is used. The general features are quite similar to Talbot and Grewal's theoretical predictions.

A more detailed study of the structure of a shock in a partially ionized gas flow by using different techniques (Thompson scattering, etc.) is under way.

B. Aperture Enlargement and Interception in Ion Engines

T. D. Masek

1. Introduction

The proper design of grids for electron bombardment ion engines requires a knowledge of the grid-failure mechanism and the establishment of the criteria for failure. Under normal circumstances, the primary mechanism for failure is erosion of grid surfaces resulting from high-velocity ion impingement. Such impingement arises: (1) from direct interception of a portion of the ion beam due to improper focusing, and (2) from charge-exchange ions produced by ion-neutral atom collisions in the accelerating region. A number of studies (Refs. 15–17) have indicated that direct interception can be almost completely eliminated by proper electrode design and that the major contribution to interception will be due to charge exchange (CE).

Generally, grid failure is considered to have occurred when erosion, warping, or any other such grid process, causes engine performance to seriously deviate from normal. In estimating engine lifetime from grid erosion data, one of the most important considerations is the question of whether such effects are nonlinear, i.e., whether the erosion rate increases with the amount of erosion.

The erosion caused by CE ions has been found to be focused mainly onto the grid between apertures. However a portion of the CE may erode the apertures radially. If this small amount of CE erosion causes ion beam defocusing, direct interception may begin, rapidly accelerating the erosion process until a runaway condition exists. An experiment was set up to determine the effect of accelerating grid aperture enlargement on ion interception. The gross effects of enlargement were considered most important to support concurrent work (Ref. 15), and precision measurements were not attempted. A second objective of the experiment was to make sputtering measurements on several possible grid materials using the ion beam to bombard the material. The sputtering rate of graphite was of primary interest, but copper and molybdenum were tested for comparison.

2. Accelerating Grid Interception

The measurements of interception ion current were taken with the device shown schematically in Fig. 5. Mercury vapor, fed to the discharge chamber, is ionized by electrons produced by a tungsten filament, forming a plasma. The plasma serves as an ion source from which ions are extracted through the accelerating system. An accel-decel system was used to simulate engine operation. The figure also shows the target for collecting ions with a biased grid to suppress secondary electron emission. The sputtering sample indicated was used to take sputtering rate data.

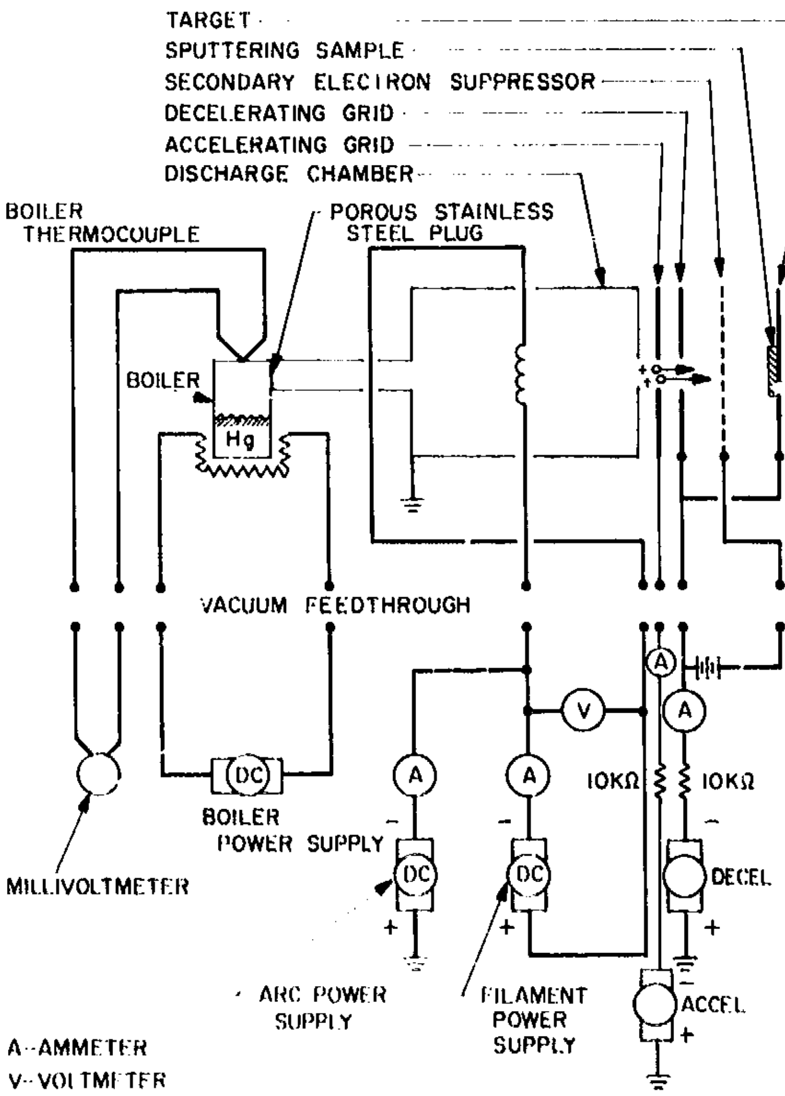


Fig. 5. Schematic diagram of an ion accelerator

The accelerating grid was so constructed that disks with different aperture diameters could be easily interchanged. The accel aperture was aligned by centering about a drill bit shaft placed through the discharge chamber aperture and the decel aperture. When changing accel apertures, only the target grid had to be completely disassembled from the chamber. All grids were supported with self-shielding boron nitride insulators mounted to the discharge chamber.

The interception data as a function of accel aperture diameter and accel-decel ratio are presented in Figs. 6–9.

In all cases, readings were taken for a series of accel-decel (A/D) ratios at constant aperture, rather than a series of apertures for constant accel-decel. This was of course because changing aperture size required shutting the system down. The decel aperture diameter remained constant. The following operating parameters were used:

Discharge chamber aperture diameter	0.200 in.
Decel aperture diameter	0.300 in.
Accel aperture diameter	0.197–0.285 in.
Beam current	20 ma
Chamber-accel grid spacing	0.375 in.
Accel voltage	6.0 kv
Decel voltage	1.0–6.0 kv
Secondary electron suppression grid voltage	200 v

It was impossible to maintain the conditions of the experiment constant through the runs. It was decided the most important parameter to maintain constant was the beam current, and the arc current and voltage were varied to accomplish this. There was no possibility of measuring the neutral efflux; it must be expected that this varied in a random manner from run to run. Because of its importance in the charge-exchange process, these

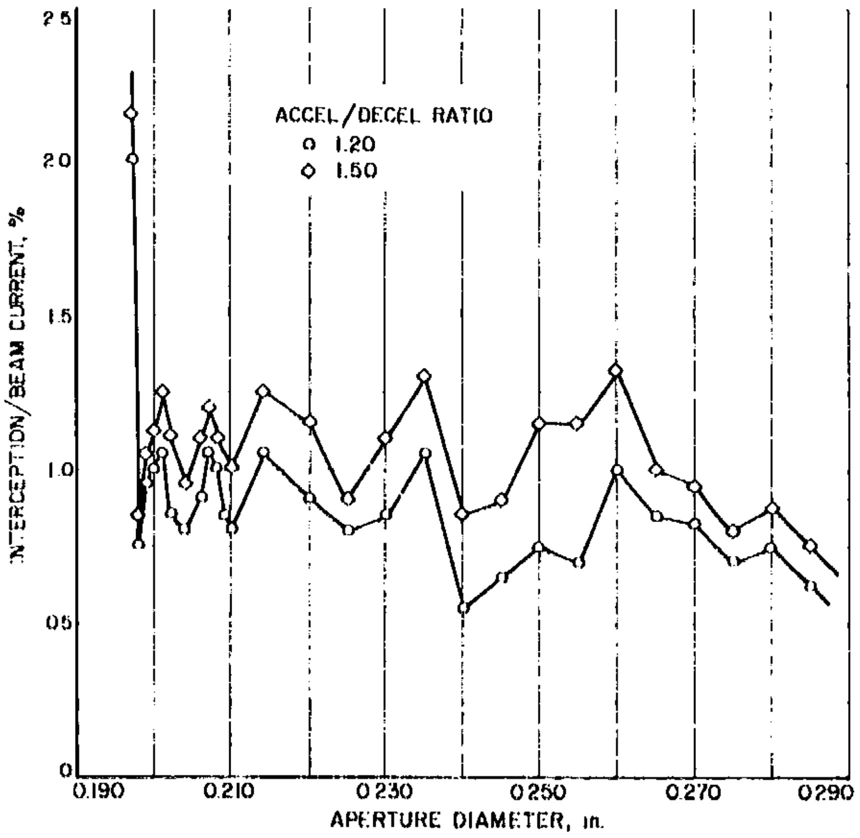


Fig. 6. Interception on accel grid for A/D ratios of 1.033 and 1.110

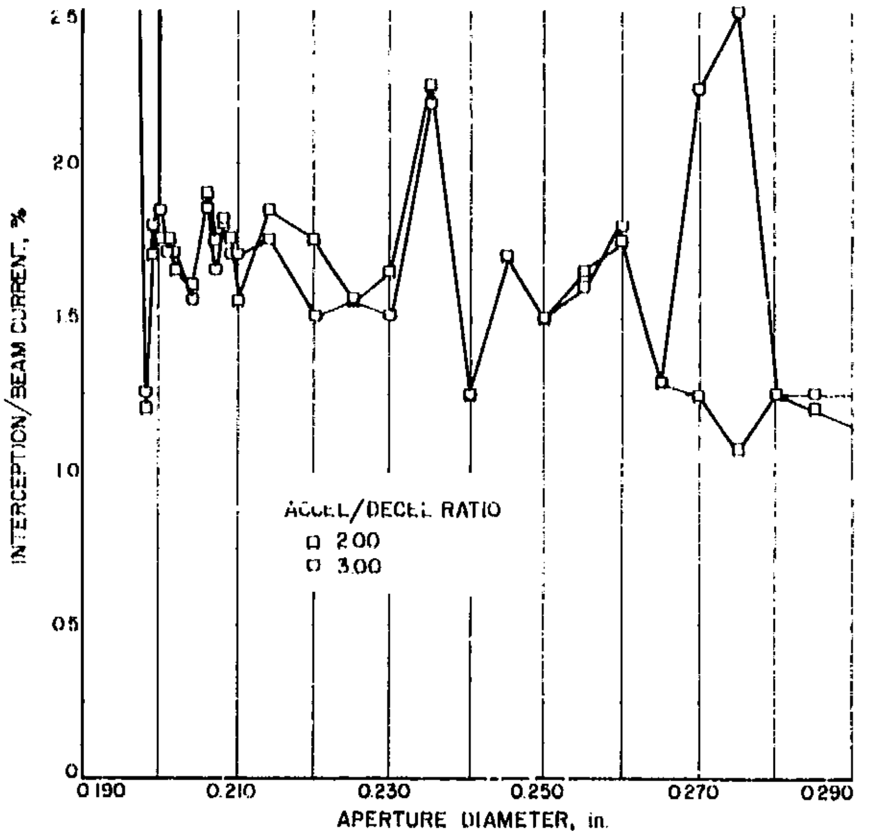


Fig. 8. Interception on accel grid for A/D ratios of 2.00 and 3.00

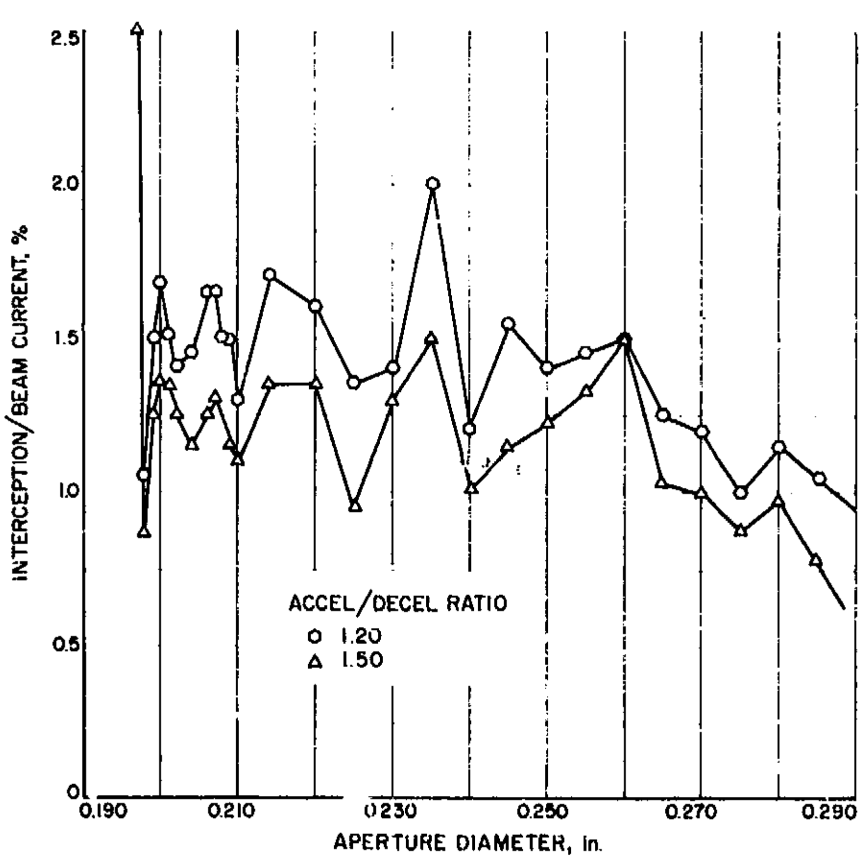


Fig. 7. Interception on accel grid for A/D ratios of 1.20 and 1.50

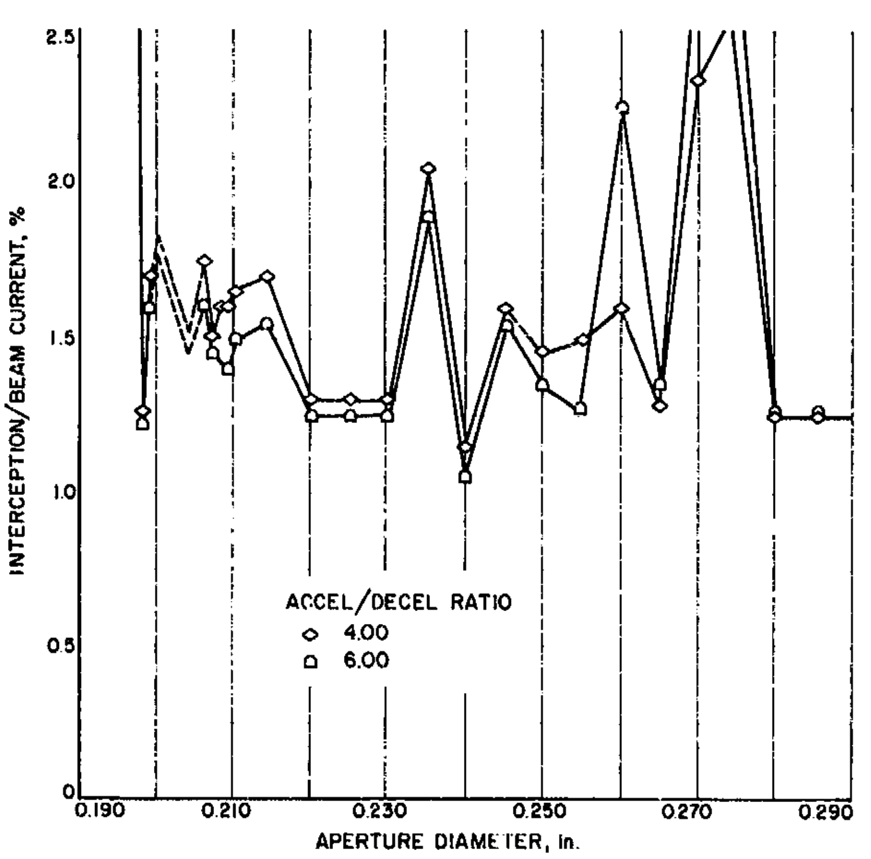


Fig. 9. Interception on accel grid for A/D ratios of 4.00 and 6.00

random variations in neutral efflux probably are responsible for the very ragged appearance of the data. At high A/D ratios and aperture diameter, interception became unstable, and data for these conditions are inaccurate.

Despite these rather large variations, the data does indicate a variation of the interception with aperture size, and that this is a generally downward trend as aperture size increases. The figures also indicate that the interception rate is maximum for A/D ratios from 2.0 to 3.0. Electron bombardment engines typically operate from 1.15 to 4.0 A/D, depending on the desired specific impulse. Interception rates for engines are also typically in the range measured (Refs. 15–17) (i.e., 0.5 to 2.5%), and the data from Ref. 16 indicates approximately the same variation with A/D.

A major consideration in interpreting the data is the origin of the interception. Calculations of charge exchange accurate enough to compare with the experiment are impossible at this time, due to the inaccuracy of charge exchange across sections. It does appear from the data that charge exchange was a significant portion of the interception. Increasing aperture diameter would decrease the ion beam density and hence would reduce the number of charge-exchange ions at constant beam current. This could account for the general downward trend with increasing aperture size.

More interesting is the variation of interception with A/D ratio. Figs. 6–9 indicate that interception increased to a maximum at A/D ratios of 2.0 to 3.0 and then dropped again. A possible explanation is as follows: At an A/D of unity, the ions enter an essentially field-free region after passing the accel grid. CE ions then would not be accelerated back to the accel grid and CE interception would be expected to be a minimum. As the decel grid becomes less negative (increasing A/D ratio), beam ions experience a decelerating force and CE ions are attracted toward the accel grid. The lower ion velocity increases the charge exchange cross section resulting in higher interception. As the beam is slowed further, it begins to diverge due to internal space charge. This effect becomes more pronounced at lower beam velocities as radially directed forces become large. As the beam diverges, the current density decreases, causing the number of charge-exchange ions to decrease. Thus, as the A/D ratio increases, the accel interception first increases until the beam space charge divergence becomes significant; it then decreases. A portion of the diverging beam however will intercept the decel grid. Decel interception could not be separately measured, since the decel grid was connected electrically to the target grid.

3. *Sputtering Rates*

The experiments reported in Ref. 15 used copper grids because of copper's high sputtering rate (Refs. 19–21). A material with a high sputtering rate was needed in order to obtain sputtering data in a short period of time. Graphite was also considered as a grid material, but, as

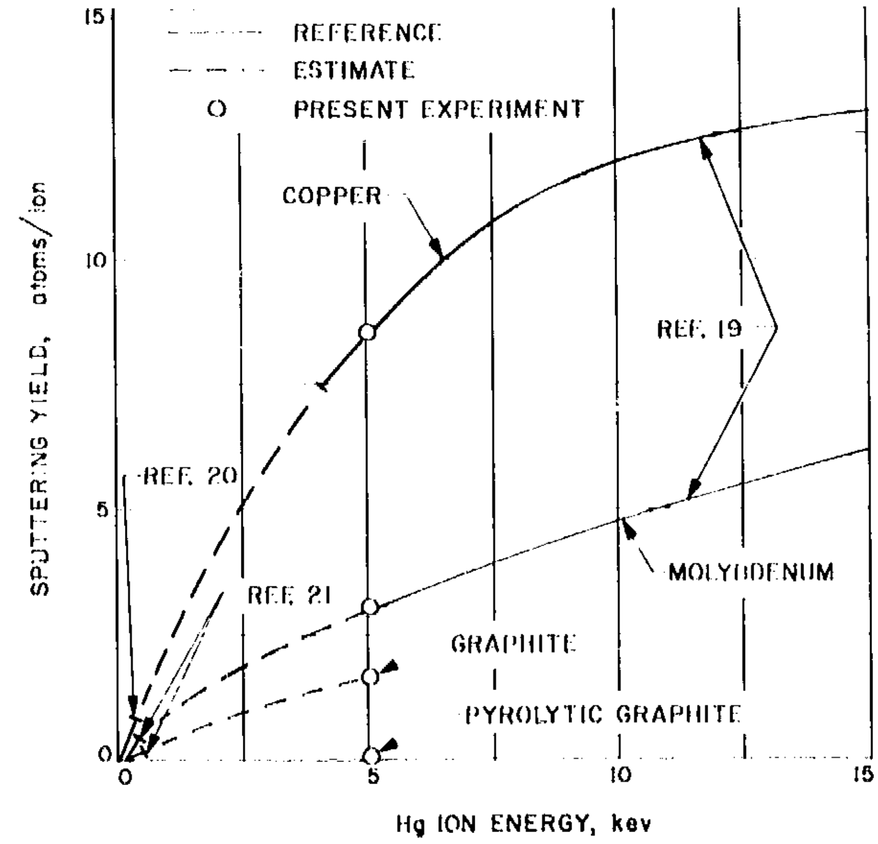


Fig. 10. Sputter yield (atoms/incident ion) for copper, molybdenum, graphite and pyrolytic graphite by bombarding Hg ions at 5 kev

little sputtering data was available, an auxiliary experiment was set up to evaluate its sputtering properties.

The previously described apparatus was used to make sputtering measurements on graphite, pyrolitic graphite, copper, and molybdenum. Since data for copper and molybdenum were published, these were used for comparison. The data points obtained are shown in Fig. 10. The correspondence of the results on copper and molybdenum with published data were sufficient to lend confidence to the rates determined from the other materials. Only the relative magnitudes of the sputtering rates were desired, therefore only a few points were taken. It was apparent that graphite and especially pyrolitic graphite, because of their low sputtering rates, would be unsuitable for the particular application in mind. Microphotographs were made of the sputtering surfaces for reference in analyzing grid erosion. Microphotographs at 325X are shown in Figs. 11–14 with unsputtered surfaces for comparison.

4. *Conclusions*

The major objective of the study, to determine if aperture enlargement affects interception in a nonlinear manner, was achieved. The study shows fairly con-

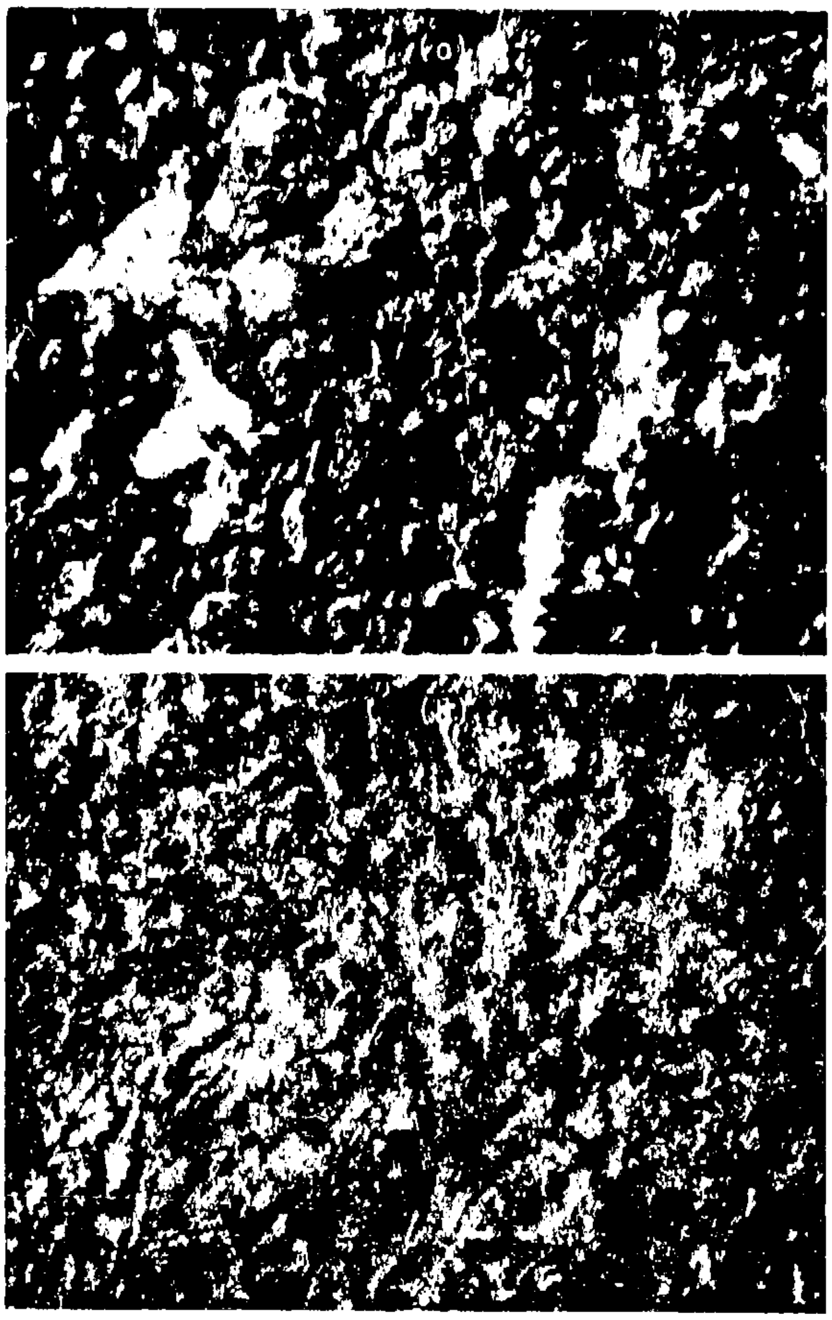


Fig. 11. Microphotographs (325X) (a) copper sputtered by 5 kev Hg ions. (b) unsputtered copper

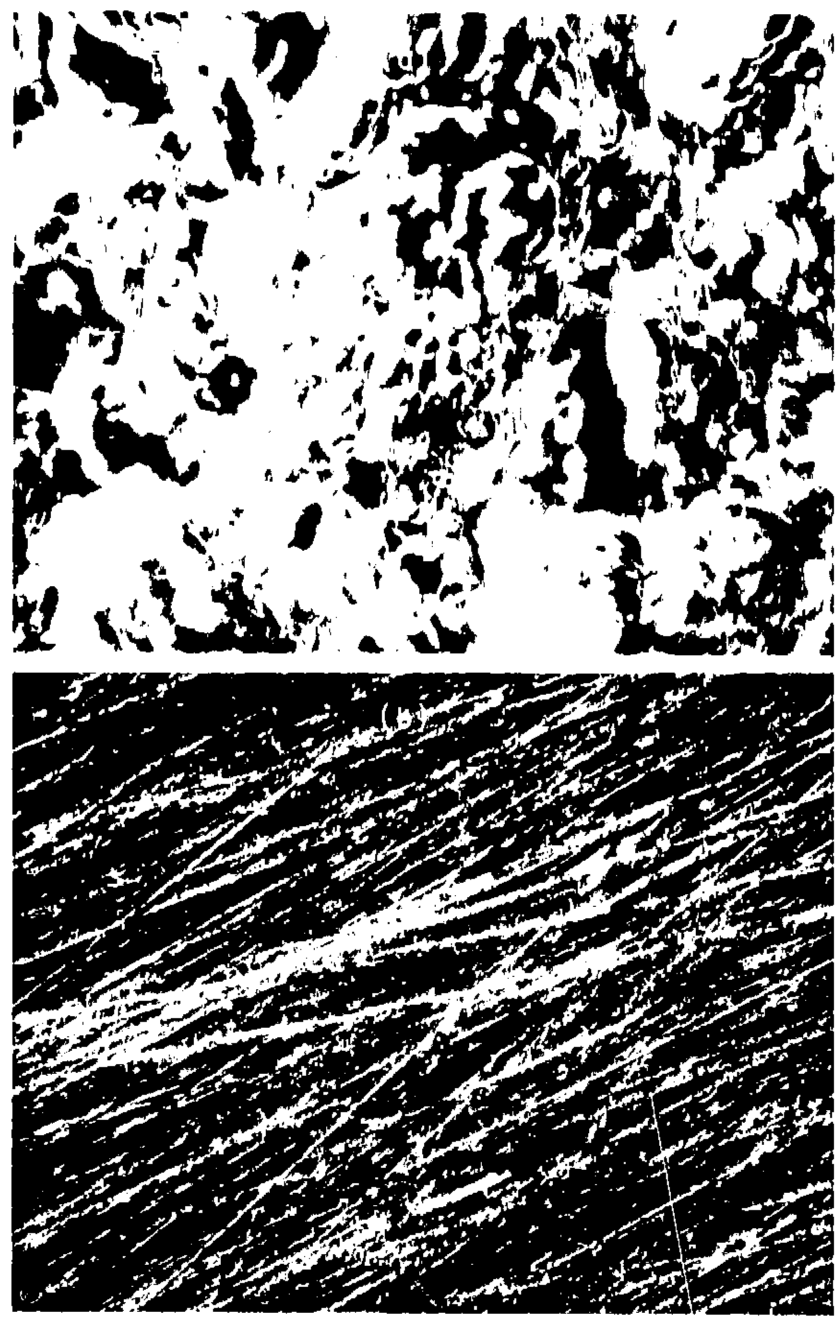


Fig. 12. Microphotographs (325X) (a) molybdenum sputtered by 5 kev Hg ions. (b) unsputtered molybdenum

clusively that interception will remain bounded for enlargements of at least 40%. Based on the data obtained in Ref. 15, it can be shown that normal grids would be worn completely away between apertures by the time apertures had been enlarged by about 15%. This implies that low interception rates will persist as long as the grid is intact. This has been verified by a recent 760-hr test of a surface contact ion engine.[1] In that test the interception remained normal up to the time of failure, which was caused by the complete wearing through of the webbing between grid apertures.

The study also indicates the desirability of low A/D operation, and that although high A/D ratios may decrease interception to the accel grid compared to the maximum encountered at A/D's of 2 to 3, increased interception to the decel grid may result.

A number of secondary conclusions have also resulted from this investigation. The study has led to a better understanding of grid design and beam-focusing problems and indicates that precision control of aperture diameter in fabrication of grids is unnecessary. This allows tolerances to be relaxed and reduces grid fabrication costs. In addition, the results show that gross sputtering-rate measurements can be made on various materials relatively simply, and the results will be applicable to grid design.

[1]Eilenberg, S. L., private communication.

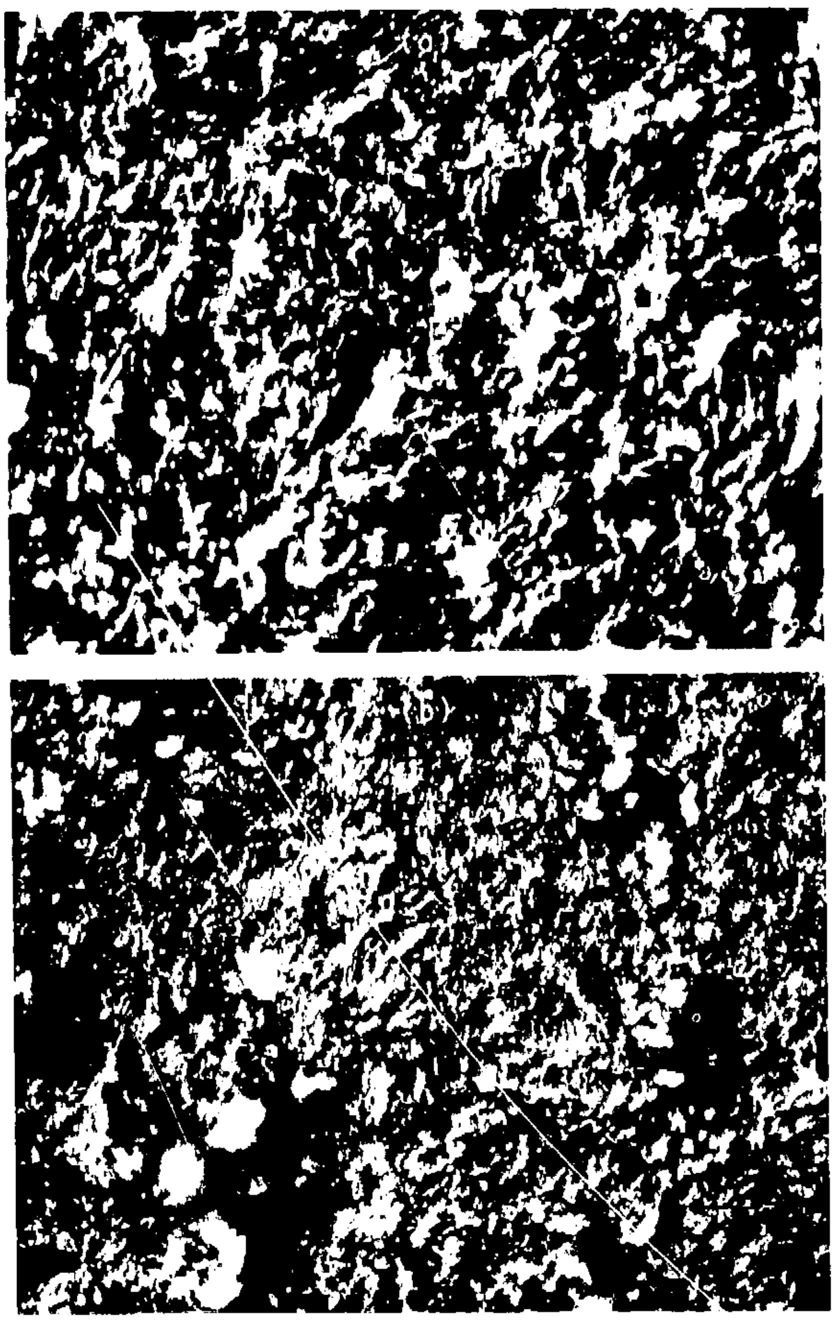


Fig. 13. Microphotographs (325X) (a) graphite sputtered by 5 kev Hg 10 ions. (b) unsputtered graphite

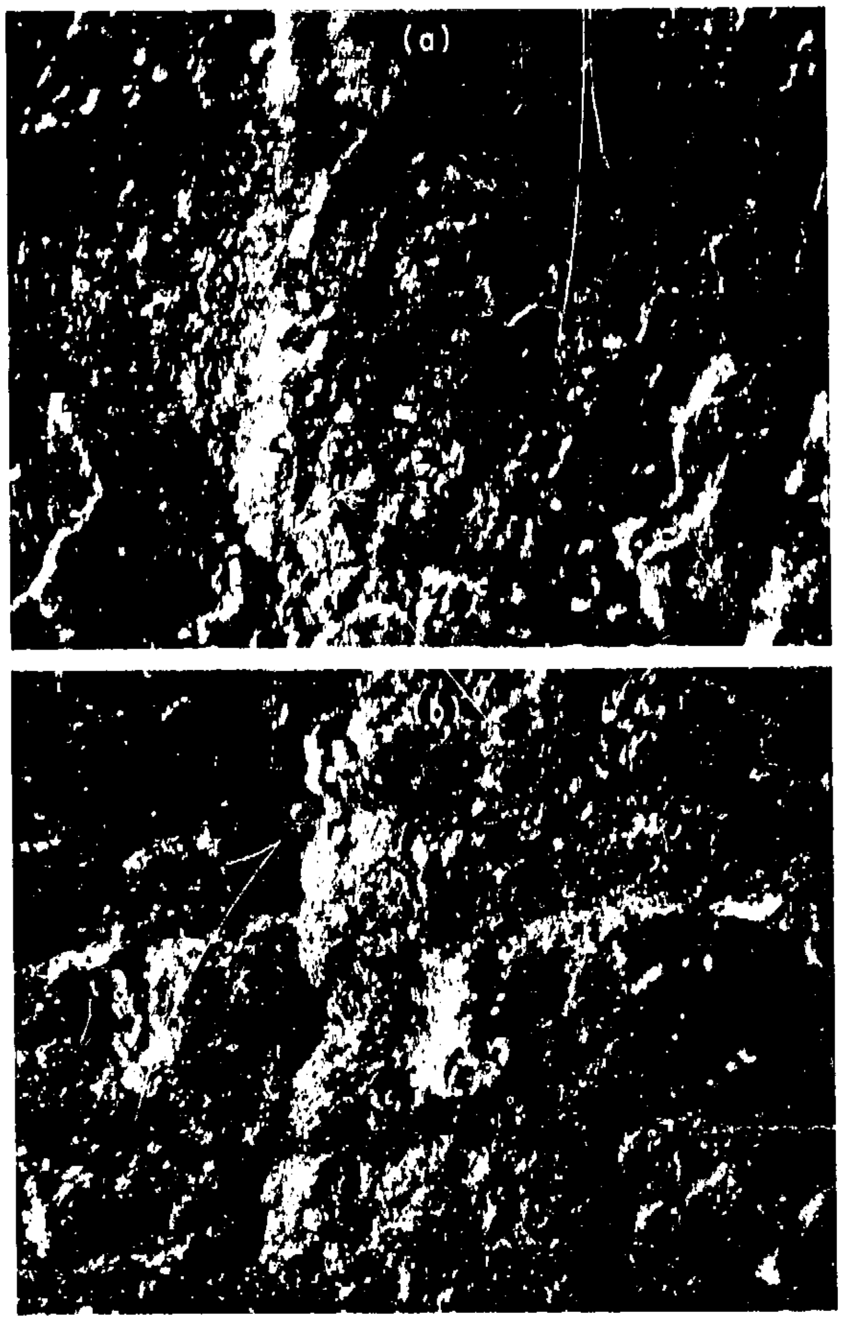


Fig. 14. Microphotographs (325X) (a) pyrolitic graphite supttered by 5 kev Hg 10 ions. (b) unsputtered pyrolitic graphite

C. Liquid MHD Power Conversion

D. Elliott, D. Cerini, and L. Hays

The long lifetimes required of electric-propulsion powerplants make nonrotating cycles attractive. A nonrotating powerplant under investigation at JPL is the liquid magnetohydrodynamic system shown schematically in Fig. 15. In this cycle a fluid, such as cesium, circulates in the vapor loop and causes a liquid metal, such as lithium, to circulate through an MHD generator in the liquid loop. The cesium leaves the radiator as condensate, flows through an EM pump to the mixer, vaporizes on contact with the lithium, atomizes and accelerates the lithium in the nozzle, separates from the lithium in the separator, and returns to the radiator. The lithium leaves the separator at high velocity (typically 500 ft/sec), decelerates through the production of electric power in the MHD generator, and leaves the generator with sufficient velocity (typically 300 ft/sec) to return through a diffuser to the reactor where the lithium is reheated.

Results of nozzle, separator, and diffuser investigations, and estimates of cycle performance, were presented in Ref. 22, and DC generator experiments were reported

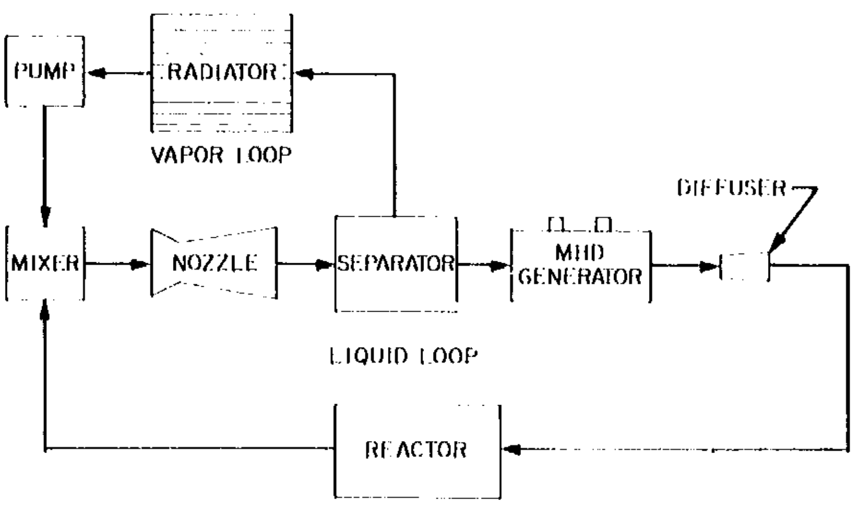


Fig. 15. Liquid MHD power conversion cycle

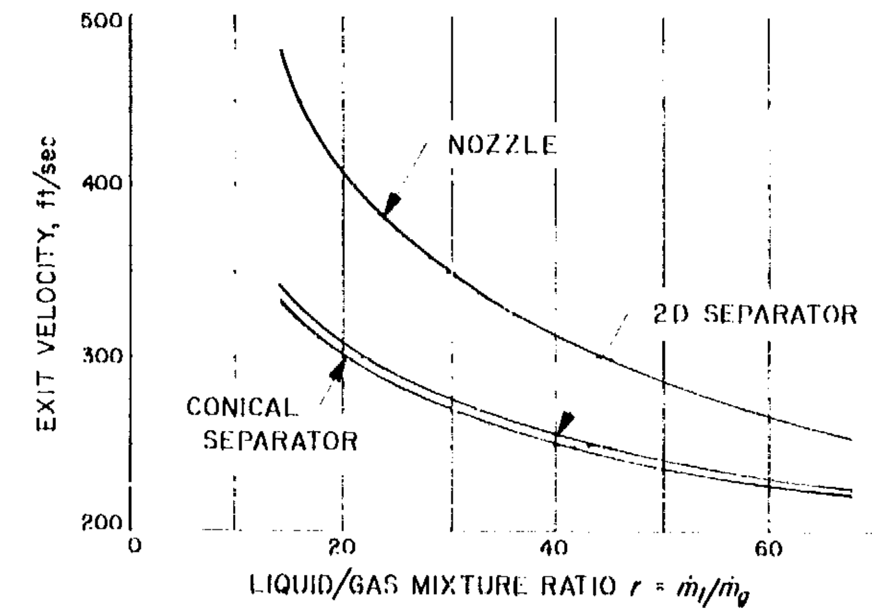


Fig. 17. Comparison of exit velocities of conical and two-dimensional separators

in Ref. 23. Further electrical and hydraulic investigations were discussed in Refs. 24–28 and are continuing.

1. Separator Experiments

The two-dimensional separator described in Ref. 24 was modified by removal of the shaded portion of the capture slot walls as shown in Fig. 16. This modification permitted direct thrust measurement of the flow entering the capture slot and thus calculation of the flow velocity leaving the separator. The variation of separator exit velocity with mixture ratio is shown in Fig. 17 for the conical and two-dimensional separators. This figure supersedes Fig. 10 of Ref. 24 which showed the two-dimensional separator to have a lower exit velocity than the conical separator; this result was incorrect and resulted from an instrumentation error.

In addition, a second two-dimensional separator surface was tested which was flat and inclined at 12 deg until the final curvature preceding the capture slot, whereas the original separator surface tested was flat and inclined at 10 deg for only 10 in., followed by a curved surface of 95-in. radius of curvature. The exit velocity was 2 to 5% less than with the curved separator, and the liquid loss from the primary capture slot was doubled, showing that curvature of the main portion of the separator is necessary.

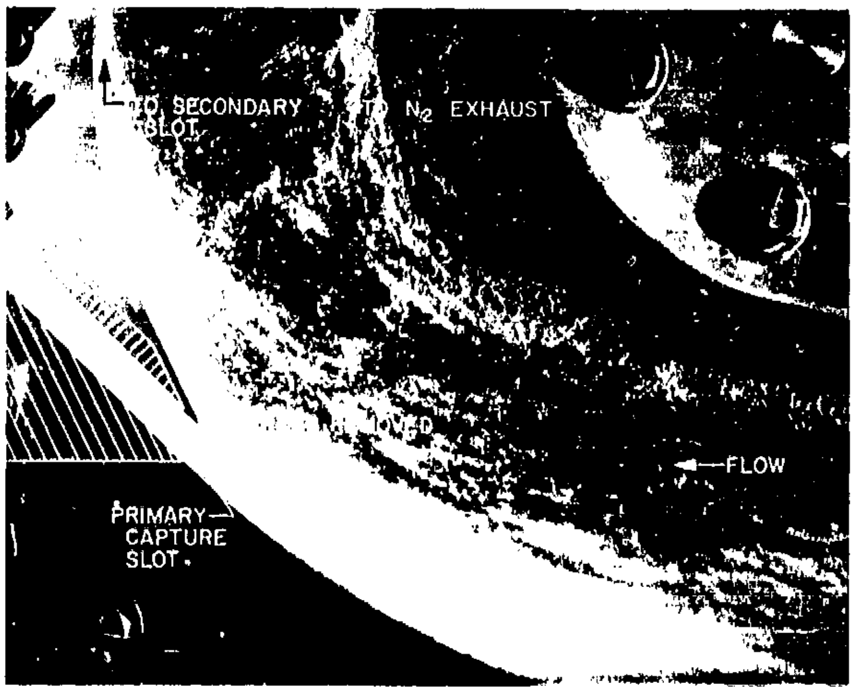


Fig. 16. Modified capture slot of two-dimensional separator

2. Generator Friction Reduction

The attainable generator efficiency would be increased if fluid friction in the generator channel leading from the capture slot could be reduced. An experiment to reduce the friction by injection of gas between the fluid and wall was performed, using the separator modification shown in Fig. 18.

The conical separator was modified by the addition of a movable inner capture slot wall which permitted the injection of nitrogen between the liquid stream and the inner wall (injection of nitrogen between the liquid stream and the outer wall was obtained automatically from the nitrogen flow along the separator). The inner and outer walls were moved independently, and thrust measurements were taken over the range $W_1 = 0.1$ to 0.4 in. and $W_2 = 0$ to 0.3 in. The results are shown in Fig. 19

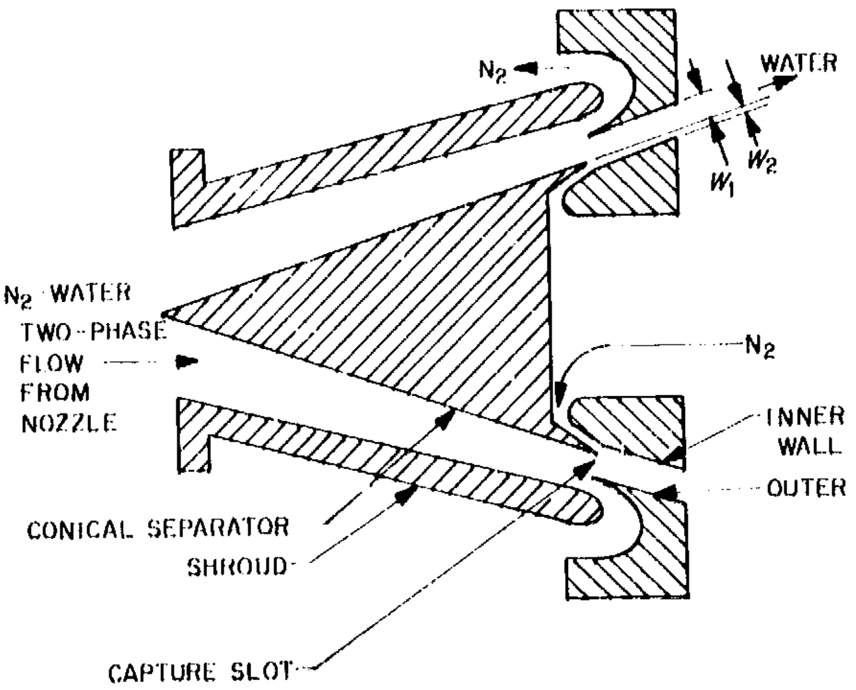


Fig. 18. Conical separator modification for boundary injection of gas

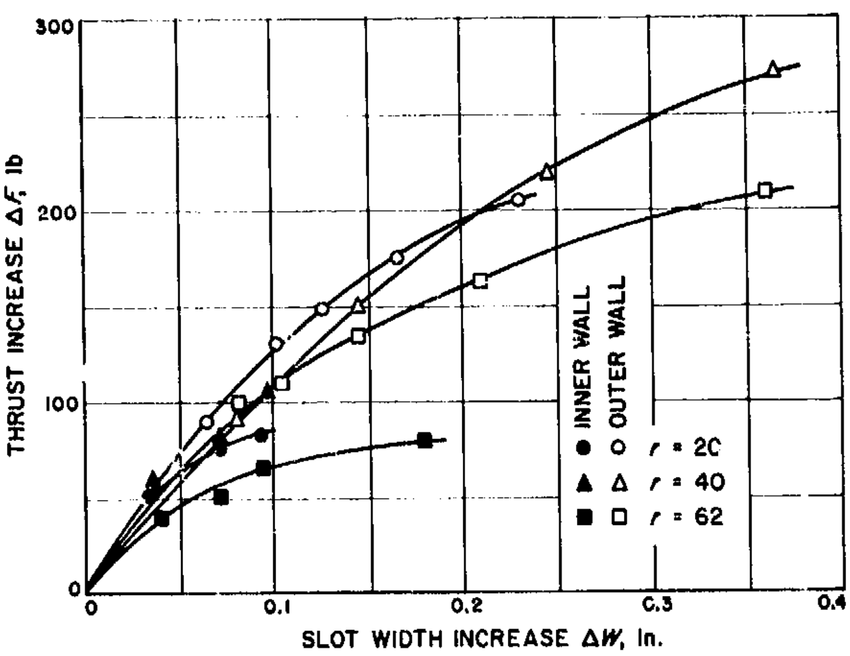


Fig. 19. Increase in thrust of liquid leaving capture slot due to slot width increase with boundary gas injection

where ΔF, for the outer wall, is the thrust increase obtained with W_1 increased from W_L to $W_L + \Delta W$, where W_L is the value of W_1 corresponding to single phase liquid flow through the capture slot. For the inner wall, ΔF is the thrust increase obtained by moving the inner capture slot wall a distance $W_2 = \Delta W$ and injecting nitrogen to maintain atmospheric pressure in the capture slot. It is seen that moving the outer wall had the largest effect.

The skin-friction coefficient C_f can be defined as the wall shear divided by the liquid dynamic pressure, and this C_f is related to the velocity ratio between slot inlet and exit by

$$C_f = \frac{2\dot{m}}{\rho V_1 A_w}\left(\frac{V_1}{V_2} - 1\right) \tag{1}$$

where $\dot{m}$ is the mass flow rate of liquid in the capture slot, A_w is the total surface area of the inner and outer walls, and ρ is the liquid density. V_1 is the flow velocity entering the capture slot (known from previous experiments) and V_2 is the flow velocity leaving the slot calculated from

$$V_2 = V_1 - \frac{\Delta F_{max} - \Delta F_T}{\dot{m}} \tag{2}$$

where ΔF_{max} is the difference in thrust between the fluid entering and leaving the slot, extrapolated to $W_1 = W_L$ at $W_2 = 0$, and ΔF_T is the total thrust gain given by Fig. 19 with both walls moved back an equal distance ΔW to obtain a total increase ΔW_T.

Fig. 20 presents the resulting calculated variation of skin-friction coefficient with slot width, with C_f normalized by the value for fully developed flow of pure liquid, C_{f_L}.

It is seen that the skin friction coefficient can be reduced by one half if the slot width is increased to three times the width required for pure liquid flow. The effect on generator efficiency is under study.

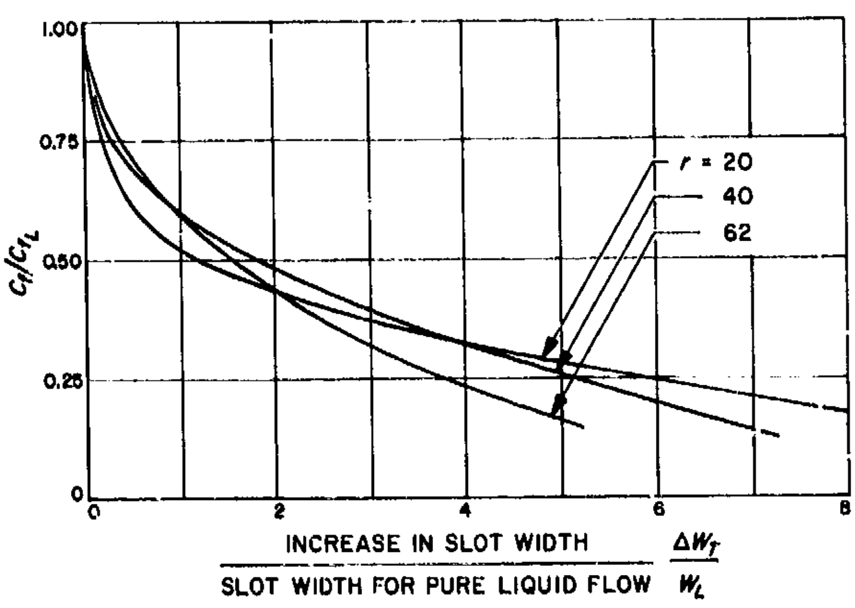


Fig. 20. Effect of slot width increase with boundary gas injection on skin-friction coefficient

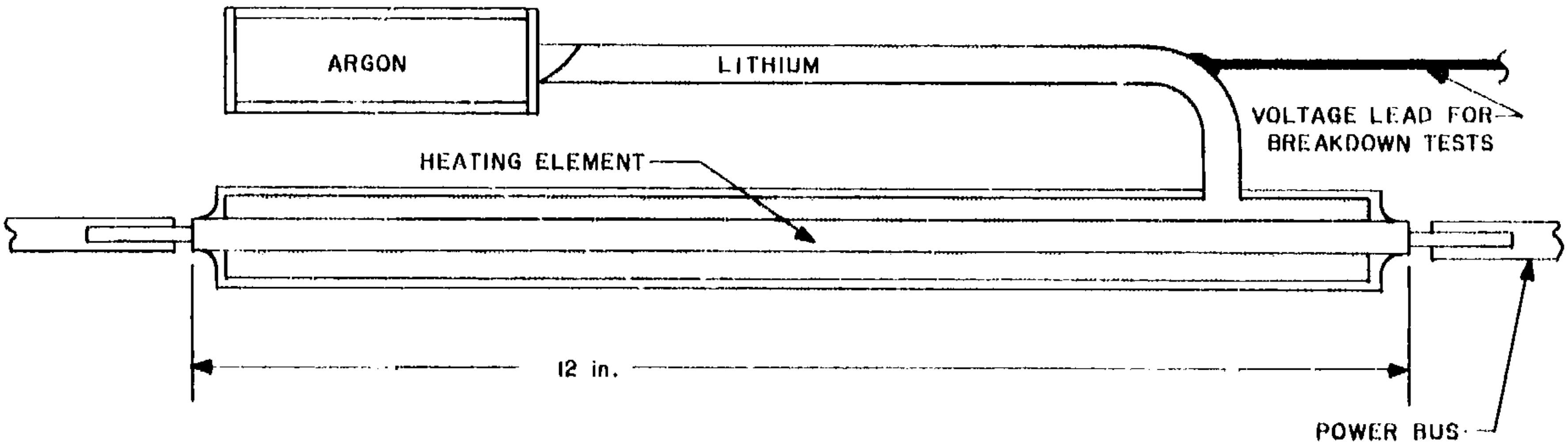


Fig. 21. Lithium heater and test capsule

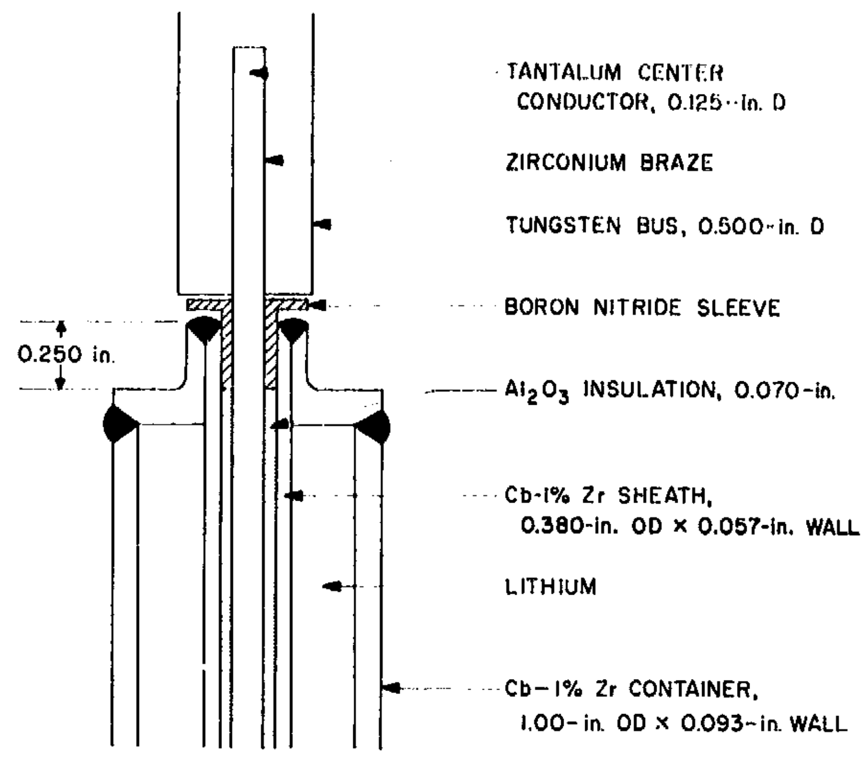


Fig. 22. End detail of heater and capsule

3. *Lithium Heater Test*

Development is proceeding on a 150-v heater suitable for heating lithium to 2000°F at high power levels (5 Mw). A high voltage device is favored over low-voltage resistance heating of the lithium, due to the greater availability of high-voltage DC power sources. The heater will be used in future testing of a 5-Mw(th) liquid MHD conversion system.

A "calrod" type of heater was selected because of its simplicity and ease of fabrication. Several test geometries were fabricated for testing in lithium at elevated temperatures under varying electrical conditions and heat fluxes. The first of these is shown schematically in Figs. 21 and 22. A tantalum rod was swaged with alumina insulation and a Cb-1%Zr outer tube. The outer tube was welded into a Cb-1%Zr container which was then filled with lithium.

The alumina insulation was recessed at the ends so that the heater sheath could be welded without degrading the ceramic. A tungsten bus was brazed to the center conductor for radiant rejection of the heat generated in the resulting bare segment of the tantalum conductor, and boron nitride inserts were provided to ensure conductor isolation from the sheath and bus.

After fabrication and lithium fill, the assembly was installed on the flange of a water-cooled vacuum chamber, Fig. 23. During operation in the chamber the heat generated in the heater was to be conducted through the lithium and container wall and transferred to the vacuum chamber wall by radiation.

Initial tests have been conducted to verify the design of the heater and the radiation cooled buses and to obtain information on the electrical resistivity and breakdown strength of the ceramic at elevated temperatures. Initial operation was entirely successful. A bulk lithium temperature of 2020°F was maintained continuously for 500 hr with a heat flux at the heater surface of 130 w/in.^{2} Low bus temperatures (~500°F), combined with the high values of center conductor temperature (~2500°F), qualitatively confirmed the validity of the radiative cooling calculations for the buses.

A value of 0.26×10^6 Ω-cm was measured for the resistivity of the alumina at its maximum temperature of

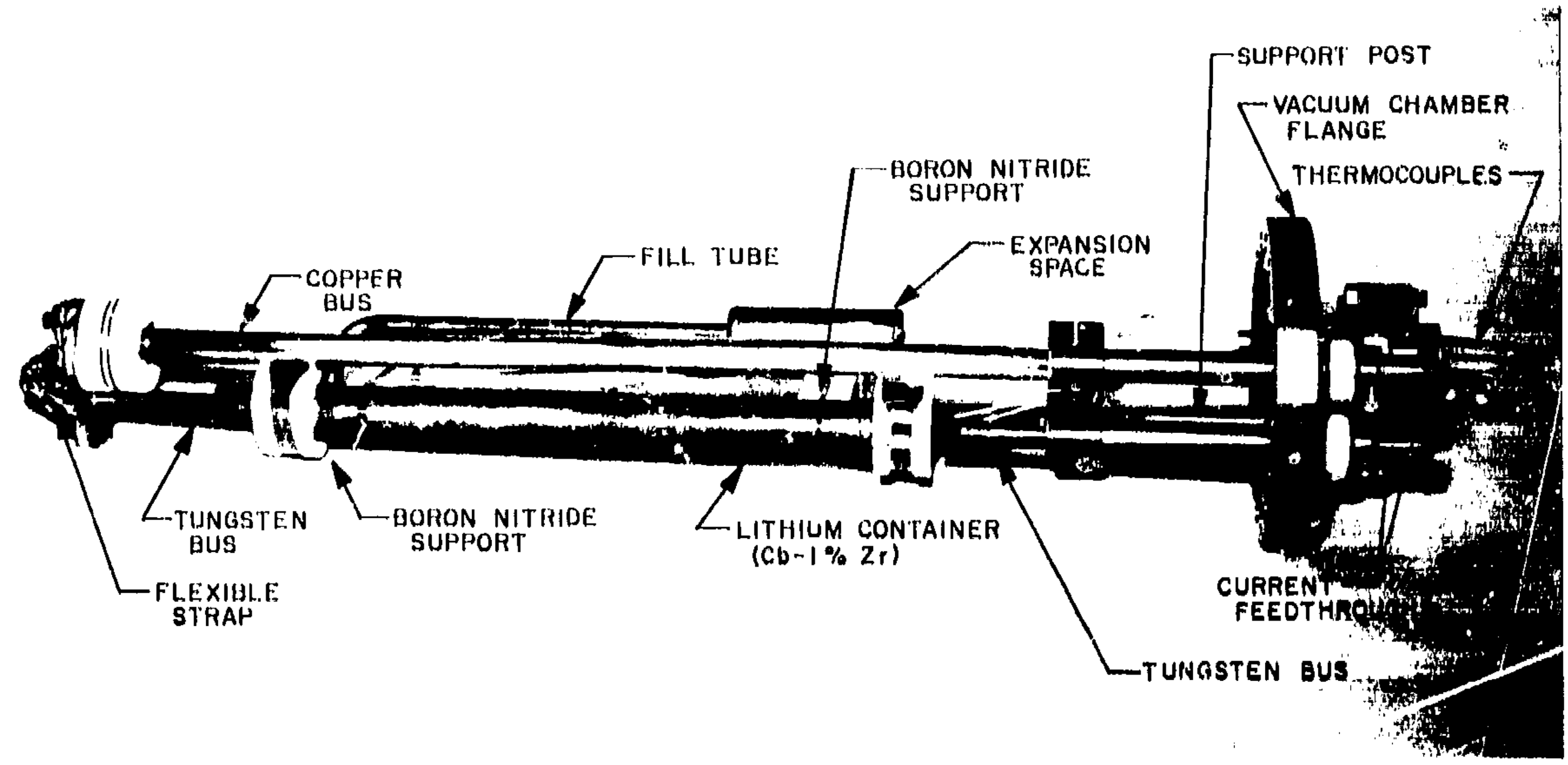


Fig. 23. Lithium heater experiment

about 2500°F as compared with the Ref. 28 value of $1 \times 10^{[illegible]}$ Ω-cm for fully dense alumina. It was found possible to sustain a voltage difference of 200 v between the sheath and center conductor of the operating temperature. A subsequent heater will have beryllia insulation and will have design parameters identical to the full-scale 5-Mw heater.

D. Variational Equations for Reactor Kinetics with Feedback Terms

H. Gronroos

1. Formulation of the Problem

A major aspect of any reactor kinetics investigation is the proper reduction of space-time equations describing the neutron and thermal balance. The purpose of the reduction is to yield equations more tractable to analytical evaluation of quantities of interest. Necessarily, the cost of the simplifications will be paid in the accuracy, and thus one criterion for the approximation processes is to find schemes likely to generate the smallest errors.

The Boltzmann neutron transport equation and the associated heat-transfer and control system equations are, in the general case, nonlinear. Linearizations and approximations employing separated variables should be done in a consistent manner that fits into the analytical techniques used for control and stability analysis. Also, for analog computer studies it is convenient to find expressions with only time as the independent variable, i.e., the proper average over the phase-space coordinates must be found. The calculus of variations coupled with dynamic programming and invariant imbedding are natural analytical tools to use in these problems.

The standard analytical techniques are: perturbation theory, adiabatic approximation, or harmonic expansions when one wants to find phase-space and time separated formulations for the kinetics problem. More recently the variational approach has been studied by several investigators. However, the connection between this technique and other above mentioned analytical methods has not been clearly demonstrated. It is the purpose of this analysis to point out the specific connections. When employing the calculus of variations one seeks a functional whose Euler-Lagrange equation yields the original equation for the neutron flux $\Phi(\bar{p}, t)$, ie.,

$$\Phi = H\Phi + S \qquad (1)$$

This functional is of the form

$$J\left[\Phi, \Phi^{+}\right] = \int_{t_0}^{t_1} \int_{\bar{p}} \Phi^{+}(\bar{p}, t)\, F\left[\Phi(\bar{p}, t), \dot{\Phi}(\bar{p}, t)\right] d\bar{p}\, dt \tag{2}$$

In the above, $\bar{p}$ stands for the phase-space coordinates, t is the time variable, H is the net production probablity operator for neutrons in the Boltzmann transport equation formulation, and S is an independent source term. Φ^{+} is an as yet undetermined weighting function. Eqs. (1) and (2) simply state that rendering the functional $J\left[\Phi, \Phi^{+}\right]$ stationary is equivalent to solving the equation for $\Phi(\bar{p},t)$. Substitution of the trial functions

$$\begin{aligned} \Phi(\bar{p}, t) &= \phi(t)\, \Psi(\bar{p}, t) \\ \Phi^{+}(\bar{p}, t) &= \phi^{+}(t)\, \Psi^{+}(\bar{p}, t) \end{aligned} \tag{3}$$

where $\phi(t)$ and $\phi^{+}(t)$ are not specified, and performing the integration over phase space yields a reduced Lagrangian

$$J\left[\phi, \phi^{+}\right] = \int_{t_0}^{t_1} \phi^{+}(t)\, F_r\left[\phi(t), \dot{\phi}(t)\right] dt \tag{4}$$

whose Euler-Lagrange equations are solved. The phase-space dependence is separated out and a simplification has been accomplished.

The significance of the above general scheme and the interpretation of the weighting function has been discussed particularly by Lewins (Refs. 29, 30 and references therein). Suffice to say that in nonlinear problems the outlined procedure required additional somewhat uncertain interpretations.

Becker and Fenech (Ref. 31) multiply Eq. (2) by an operator

$$g = H^{*} W(\bar{p}), \tag{5}$$

where H^{*} is the adjoint operator defined by

$$\int_{\bar{p}} \Phi^{*} H \Phi\, d\bar{p} = \int_{\bar{p}} \Phi H^{*} \Phi^{*}\, d\bar{p} \tag{6}$$

and $W(\bar{p})$ is an arbitrary positive weighting function, to obtain the least squares functional

$$J = \int_{t_0}^{t_1} \int_{\bar{p}} W(\dot{\Phi} - H\Phi - S)^2\, dt\, d\bar{p}. \tag{7}$$

This takes care of nonlinearities, but the transformation of the original equation introduces new boundary conditions, and information is lost in the process (Ref. 32). However, there are also many advantages in the above procedure.

The approach by Selengut (Ref. 33) consists of finding the functional

$$J[\phi] = \int_{t_0}^{t_1} F\left[\phi(t), \dot{\phi}(t)\right] dt \tag{8}$$

given the Euler-Lagrange equation

$$L[\phi] = \partial F/\partial\phi - \partial/\partial t\, \partial F/\partial\dot{\phi} = 0 \tag{9}$$

A trial function

$$\phi(t) = f(a_i, t) \tag{10}$$

with the a_i's as variational parameters may be used. As Selengut treats this problem, it may not always be possible to find a proper functional. In addition phase-space coordinates are not considered.

Starting out from different physical considerations, namely those involved in dynamic programming and invariant imbedding, one can find a functional in physically admissible cases as will be shown below. Also the separation of the phase-space dependence may be achieved from these considerations. Close similarities with Lewins' and Selengut's treatment will be noticed but a basically different philosophy will separate this treatment from the other approaches to the reduction process.

2. *Derivation of the Variational Functional*

Suppose we have an as yet unspecified functional

$$J[\phi] = \int_{a}^{T} F\left[\phi(t), \dot{\phi}(t)\right] dt \tag{11}$$

pertaining to functions of the form

$$\dot{\phi} = f^1(\phi, \tau, t), \ \phi(a) = c \tag{12}$$

$$\dot{\tau} = f^2(\phi, \tau, t), \ \tau(T) = b. \tag{13}$$

$\phi(t)$ represents the phase space averaged neutron flux and $\tau(t)$ represents the temperature and supplies the feedback mechanism. The Euler-Lagrange equation is given by Eq. (9), which is a minimum over $\phi(t)$. Consider instead that the minimum of the functional Eq. (11) is a functional of arbitrary initial conditions $\phi(a) = c$ and initial instants a, i.e.,

$$f(c, a) = \min_{\dot{\phi}} J[\phi] \tag{14}$$

which may be thought of as the "cost" of the process. For a reactor, we would say that it always behaves so that over a choice of $\dot{\phi}(t)$ it picks the one that minimizes

$f(c,a)$. This is also the starting point for Lewins in Ref. 29, but his subsequent development differs from this treatment. Using the principle of optimality (Refs. 34, 35) (i.e., dynamic programming), to relate the cost of neighboring processes

$$f(c,a) = \min_{z} \left[F(c,z)\,h + f(c + zh, a + h) + 0(h)\right] \tag{15}$$

gives in the limit as h tends to zero

$$0 = \min_{z} \left[F(c,z) + z f_c + f_a\right], \; f(c,T) = 0 \tag{16}$$

z is the choice of the slope $\dot\phi$ at the time a when the displacement is c. We find z_{min} from

$$f_c(c,a) = -\partial F(c,z)/\partial z \tag{17}$$

and putting it into Eq. (16) gives the Hamilton-Jacobi partial differential equation

$$F(c,z) + z f_c(c,a) + f_a(c,a)\Big|_{\substack{z = z_{min}\\ a < T}} = 0 \tag{18}$$

with the boundary condition

$$f(c,T) = 0\,. \tag{19}$$

Thus if $F\left[\phi(t), \dot\phi(t)\right] \rightarrow F(c,z)$ is known the cost function $f(c,a)$ can be evaluated.

The appropriate $F(c,z)$ is found utilizing Eqs. (12) and (13). Define

$$\tau(a) = r(c,a) \tag{20}$$

and following the standard invariant imbedding procedure (Refs. 35, 36) and after taking limits

$$f^1(c,r,a)\,r_c + r_a = f^2(c,r,a) \quad, a < T \qquad r(c,T) = b \tag{21}$$

and

$$f^1(c,r,a)\,s_c + s_a = 0 \quad, a < T \qquad s(c,T) = c \tag{22}$$

where

$$s(c,a) = \phi(T) \tag{23}$$

Eqs. (21) and (22) are solved to obtain $r(c,a)$ and $s(c,a)$. The obvious identification from Eq. (17)

$$z_{min} = r(c,a) \tag{24}$$

gives after differentiation of Eq. (18) with respect to c

$$\partial/\partial c\left[F(c,r) - rF_r(c,r)\right] + \partial/\partial a\left[F_r(c,r)\right] = 0. \tag{25}$$

From Eq. (25) one may find $F(c,r)$, since $r(c,a)$ is known from Eq. (21) and then after the transformations $a \rightarrow t$, $c \rightarrow \phi(t)$, $r(c,a) \rightarrow \tau(t)$ one finds the functional

$$J\left[\phi, \tau\right] = \int_a^T F\left[\phi(t), \tau(t)\right] dt\,. \tag{26}$$

This is not the same as the functional Eq. (11) and additional considerations must be added. Elimination of $\tau(t)$ in Eq. (12) leads to an equation of the form

$$\ddot\phi = f(\phi, t) \tag{27}$$

The appropriate functional is

$$J\left[\phi\right] = \int_a^T \left[\frac{1}{2}\dot\phi(t)^2 + G\left[\phi(t), t\right]\right] dt \tag{28}$$

$$f(\phi, t) = \partial G(\phi, t)/\partial\phi \tag{29}$$

and following the standard procedure

$$z_{min} = -f_c(c,a) = r(c,a) \tag{30}$$

$$G(c,a) - \frac{1}{2} f_c^2 + f_a = 0 \tag{31}$$

and therefore

$$G_c(c,a) - r\,r_c - r_a = 0 \tag{32}$$

Thus the following procedure is established:

(1) From the given time dependent systems equations obtain the invariant imbedding formulation, i.e., express the problem as an initial value problem.

(2) Obtain the solution for $r(c,a)$ from the partial differential Eq. (21).

(3) Solve for $G(c,a)$ from Eq. (32).

(4) Substitute $a \rightarrow t$, $c \rightarrow \phi(t)$ in $G(c,a)$ and write Eq. (28).

Putting in trial functions in the obtained equation corresponds to Selengut's recipe. From Eq. (32) one may also obtain directly the Euler-Lagrange Eq. (29). Notice that this procedure establishes the existence of a functional and its Euler-Lagrange equation.

3. *Introduction of the Phase-Space Coordinates*

Eq. (12) may be looked upon as the result of a phase-space averaging process

$$\int_{\bar{\rho}} \psi^{+} \dot{\Phi}\, d\bar{\rho} = \int_{\bar{\rho}} \psi^{+} \left[H\Phi + S \right] d\bar{\rho} \qquad (33)$$

where ψ^{+} is the weighting function to be determined. In fact the correct weighting in the linear case is obtained from the adjoint function (Ref. 1)

$$\dot{\Phi}^* = -H^* \Phi^* \qquad (34)$$

If one had started out with a functional

$$J\left[\Phi, \Phi^*\right] = \int_{t_0}^{t_1} \int_{\bar{\rho}} \Phi^* \left[-\dot{\Phi} + H\Phi + S \right] d\bar{\rho}\, dt \qquad (35)$$

where H is a linear operator and then substituted the trial functions,

$$\Phi = \phi(t)\, \psi(\bar{\rho}, t) \qquad (36)$$

$$\Phi^* = \phi^*(t)\, \psi^*(\bar{\rho}, t) \qquad (37)$$

and had taken the variation with respect to ϕ^* one would find

$$\int_{\bar{\rho}} \psi^* \dot{\Phi}\, d\bar{\rho} \quad d\bar{\rho}\, dt \int_{t_0}^{t_1} \int_{\bar{\rho}} \psi^* \left[H\Phi + S \right] d\bar{\rho}\, dt \qquad (38)$$

and performing the phase-space integration it would give a differential equation in t. Thus the difference between perturbation theory and the variational approach is limited to substitution of trial functions. This can as well be done directly into Eq. (33) without specific recourse to a variational approach. For the system of equations

$$\dot{\Phi} = H^1(\tau)\, \Phi + S^1 \qquad (39)$$

$$\dot{\tau} = H^2(\Phi)\, \tau + S^2 \qquad (40)$$

Substitute directly the trial functions, multiply through with the weighing function and integrate over the phase-space variables. This will give the system of Eqs. (12) and (13), where for simplicity the independent source terms were omitted.

The determination of trial functions and their adjoint of course requires insight into the problem. Usually, however, from a series of static calculations reasonable trial functions may be found. Kaplan, et al. (Ref. 37), find that the flux shape in a reactor during a transient is well described by a linear combination of the initial shape $\psi(\bar{\rho}, t_0)$ and the asymptotic tilted shape.

4. *Summary*

The above-described scheme attempts to show a consistent way of reducing nonlinear time and phase-space dependent equations. Various modifications arise depending on the nature of the problem. For the nonlinear problem a consistent variational formulation cannot be found, and the trial functions must be substituted directly into the original equations and then a weighted average taken as in perturbation theory. In the linear case the variational formulation leads to the same expressions as given by this theory. After this the further calculations follow standard dynamic programming and invariant imbedding procedures.

The actual calculation may be formidable, but the preparation of properly reduced equations is still a much less time-consuming process than attempts to solve directly the original equations, even if this could be done for a realistic system. The experience is, that an order-of-magnitude increase in the analytical effort does not yield the corresponding improvement in the accuracy.

E. Plasma Characteristics for Mercury and Cesium Bombardment Ion Engines

T. D. Masek and D. J. Kerrisk

1. *Introduction*

For some time a laboratory program to evaluate electron bombardment ion engines has been in progress. In the course of this program the characteristics of a mercury plasma were determined, and the effect of these characteristics on engine performance and lifetime evaluated. The most serious of these was the nonuniform density of the plasma, the causes of which were analyzed to determine what corrective measures could be taken. Recently, work has been directed toward the cesium bombardment engine. The mercury engine was modified to accommodate the new propellant, and several of the modifications suggested by the analysis of the mercury plasma were incorporated in the modified design. Little improvement in plasma uniformity was observed, but striking differences were noted in other plasma characteristics which suggested that the causes of nonuniformity are quite different in this engine. Some of these causes and their effects are discussed below.

2. *Experimental Setup*

Fig. 24 shows a schematic diagram of the experimental engine and its attendant instrumentation. This schematic holds for both the mercury and the cesium engine, and the type of measurements taken were the same in both cases. To measure plasma characteristics, a Langmuir probe was immersed in the plasma. This probe, adjustable in position both radially and axially from outside

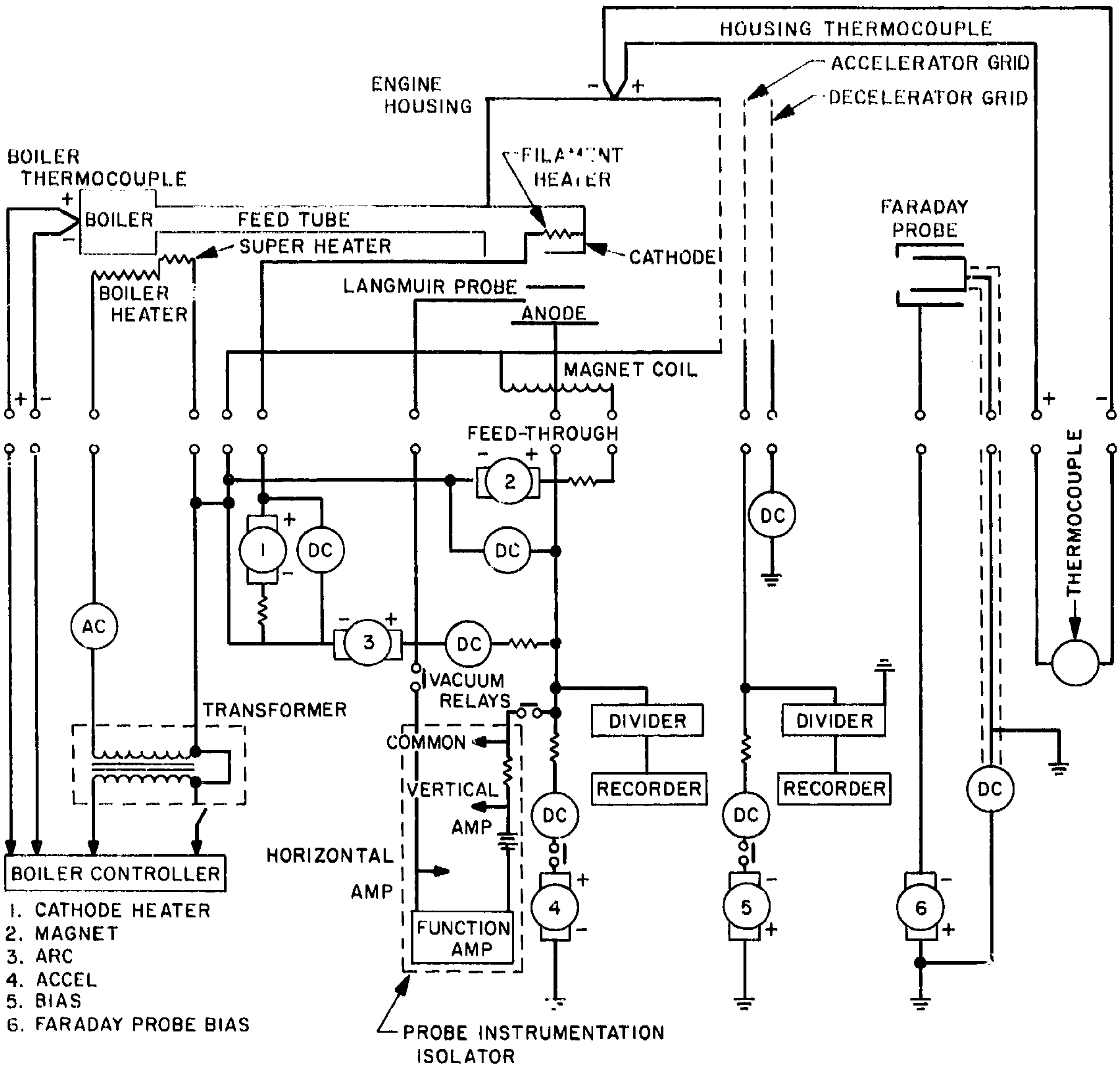


Fig. 24. Electron bombardment ion engine schematic diagram

the tank, was coupled to its instrumentation through vacuum relays. This permitted the probe, which was at high voltage, to be decoupled at will from the instrumentation without disturbing the engine operation and greatly simplified the taking of data. A typical Langmuir probe of the type used is shown in Fig. 25.

A Faraday probe, also radially and axially adjustable from outside the tank, was used to map the ion beam. The data obtained were used to correlate plasma characteristics with ion beam formation. Fig. 26 shows the cesium engine with the Faraday probe in place, and Fig. 27 shows the probe adjustment mechanisms mounted on the vacuum tank header.

3. Results

Langmuir probe theory is well known, and the probe characteristics can be used to determine ion density, plasma potential, and electron temperature, provided the electrons in the plasma have a Maxwellian energy distribution. In such case, the probe characteristic when plotted semilogarithmically has the shape shown in Fig. 28 (typical of cesium plasma measurements) where the knee of the curve gives the plasma potential and the electron density, and the slope of the straight line to the left of the knee is the reciprocal of the electron temperature. As explained in Ref. 38, this theory can be extended to include a plasma where a monoenergetic electron distribution with random velocity vectors is superimposed on a Maxwellian distribution. In this case the semilog plot has the shape shown in Fig. 29 (typical of mercury plasma) but can be corrected to the normal Maxwellian plot as shown, by subtracting the contribution of the monoenergetic group. In the plasmas under discussion here such a group could be expected, being those electrons (emitted from the cathode and having fallen through the cathode sheath drop) which have not yet given their energy to the plasma. This group is hereafter referred to as "primaries."

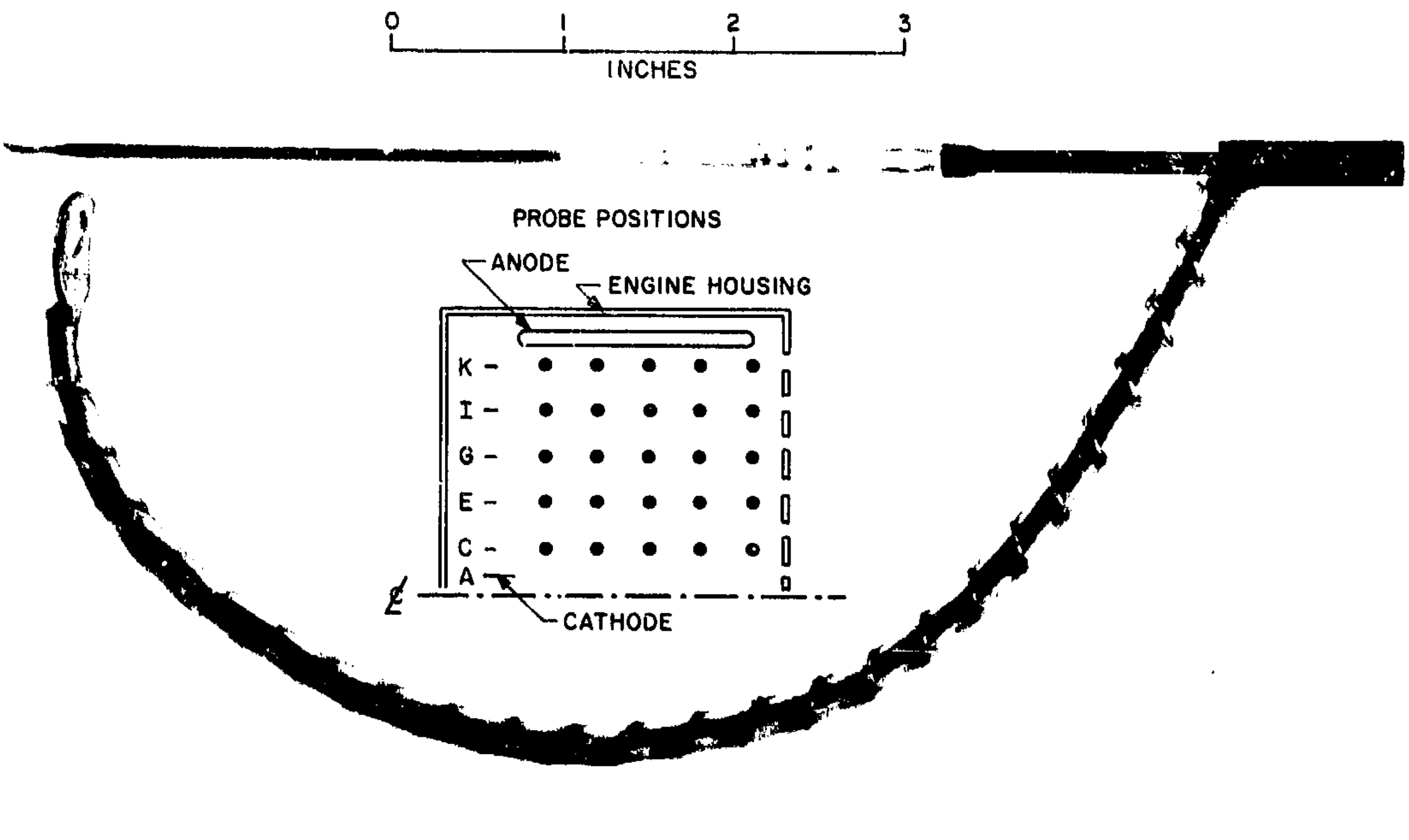


Fig. 25. Typical Langmuir probe (insert indicates probe positions)

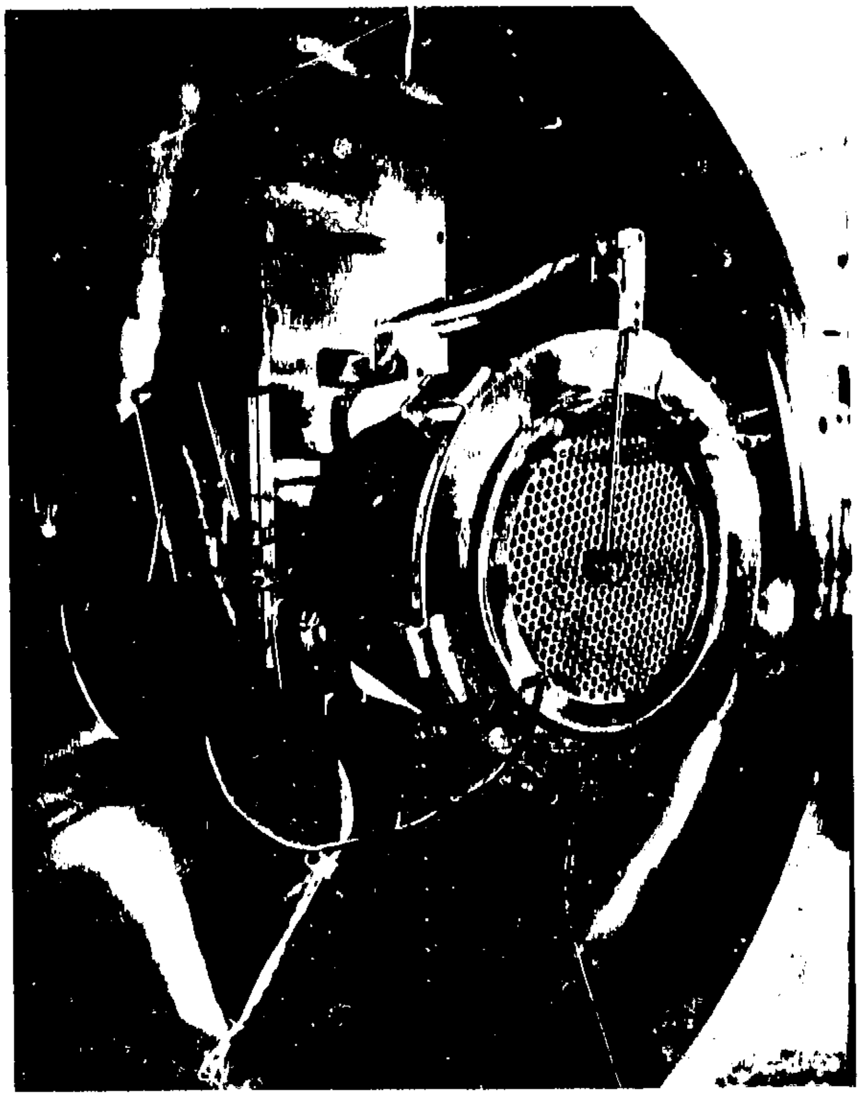


Fig. 26. Cesium electron bombardment ion engine with Faraday probe mounted

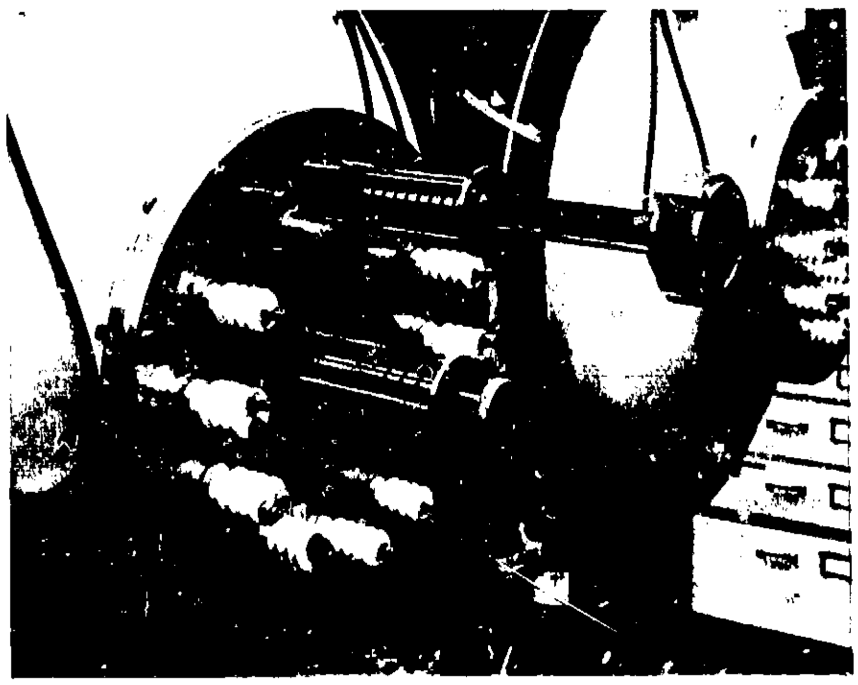


Fig. 27. Probe positioning mechanisms

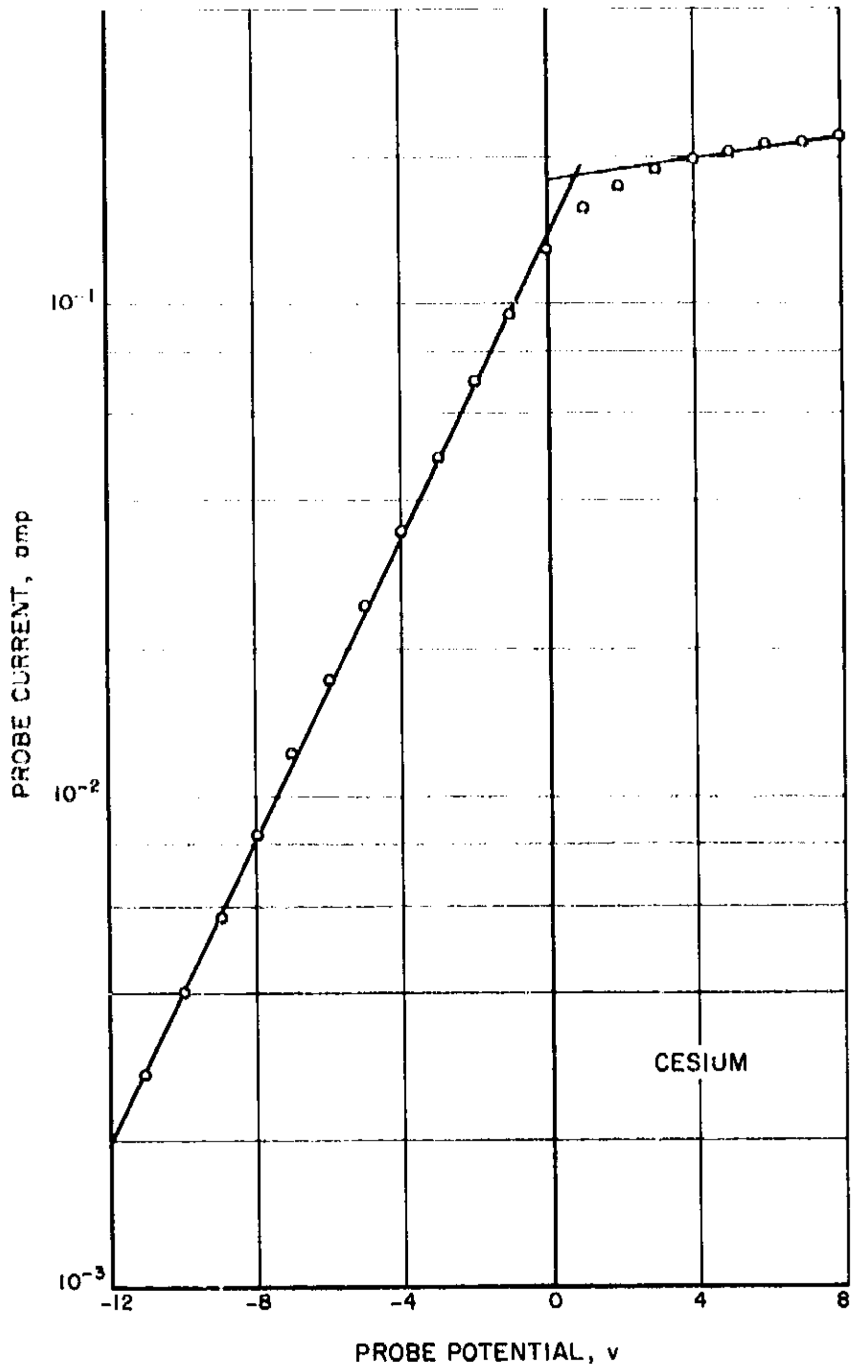


Fig. 28. Typical Langmuir probe semilog plot for cesium plasma

Table 1. Engine operating conditions

Parameter	Mercury	Cesium
Beam current, ma	210	220
Mass flowrate, g/hr	1.9	1.3
Accel voltage, kv	3.5	3.2
Bias voltage, kv	2.0	0.96
Arc voltage, v	45	14.5
Arc current, amp	3.0	7.8
Magnet current, amp	7.5	3.6
Power efficiency, %	70	74
Propellant efficiency, %	85	85

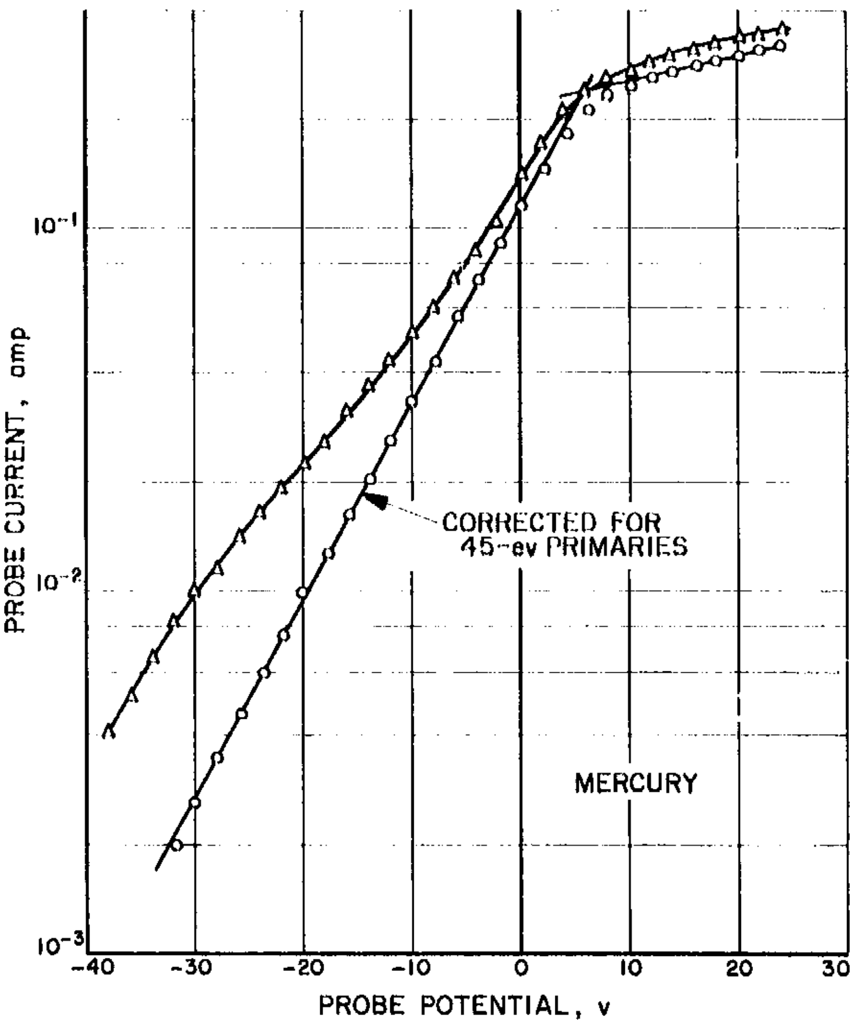


Fig. 29. Typical Langmuir probe semilog plot for mercury plasma

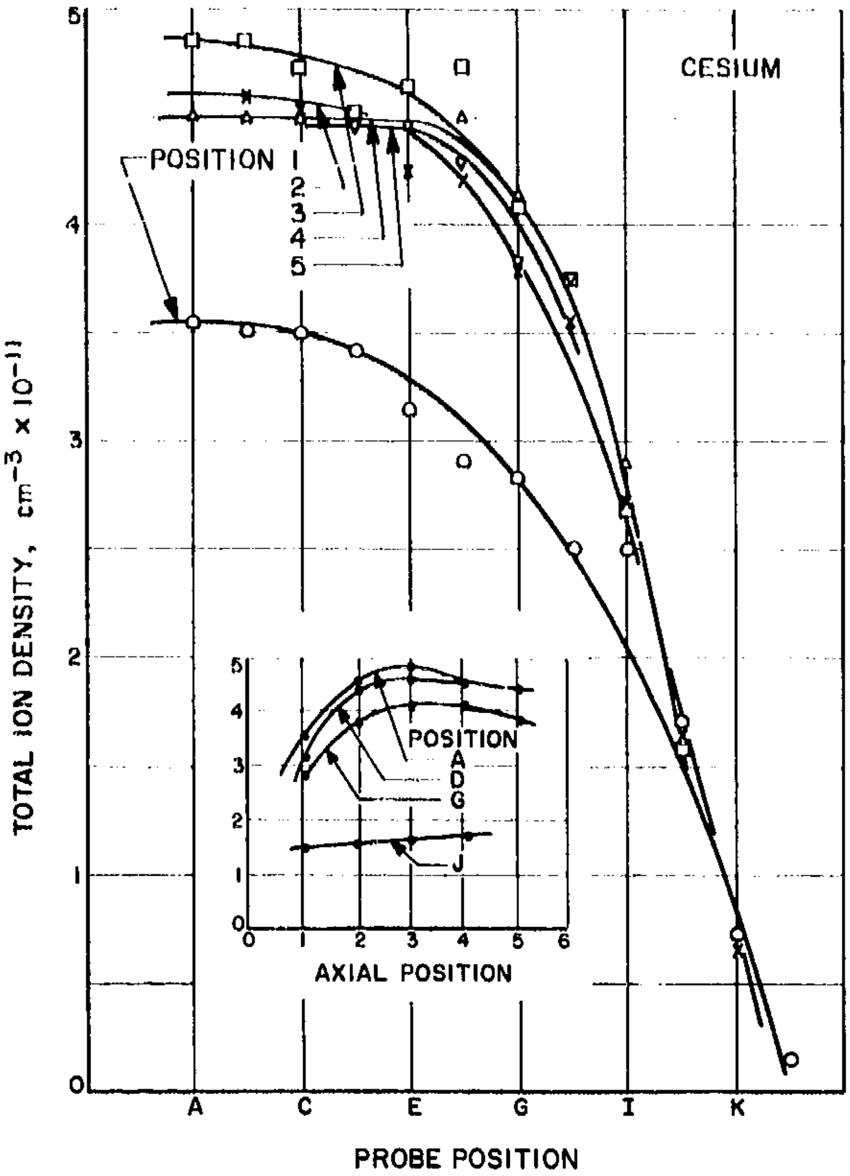


Fig. 30. Typical ion density distribution for cesium

Figs. 30–37 give the observed plasma ion density distributions, primary electron distributions, primary energy distribution, plasma potentials, and Maxwell electron energy. It will be noted that while the ion density distributions are quite similar, the other characteristic values of the plasma are quite different, and the primary electron distribution is completely absent in cesium. The conditions for the two runs for which these data were taken are given in Table 1 and are typical of a large number of runs for both propellants.

4. Discussion

In an engine of this type, ionization is produced by the collision of high-energy electrons with neutral gas particles. Because electron mean free paths are extremely long in gases of this density (10^{11} to 10^{12} particles/cm^3), an axial magnetic field is used effectively to shorten the mean free path to the electron cyclotron radius. Both primary and Maxwellian electrons can produce ionization, but while all of the former group have sufficient energy, only those in the high energy tail of the Maxwell group (i.e., those whose energy is greater than the ionization potential) have ionization capability. The arc voltages shown in Table 1 correspond approximately to the maximum ionization cross-sections for each propellant and should guarantee maximum utilization of the primaries in producing ionization.

The relative effectiveness of Maxwell electrons in producing ionization may be calculated from the ionization probability

$$P_i = \int_0^\infty F(E)\, \sigma_i\,(E)\, E^{1/2}\, d\,E \qquad (1)$$

where $F(E)$ is the electron energy distribution function, and $\sigma_i(E)$ is the ionization cross-section. For mean electron energies of 1.5 ev for cesium and 5 ev for mercury, the ratio of P_i (Hg) to P_i (Cs) is approximately 8, i.e., the Maxwellian electrons could be expected to be about eight times as effective in producing ionization in mercury as they are in cesium.

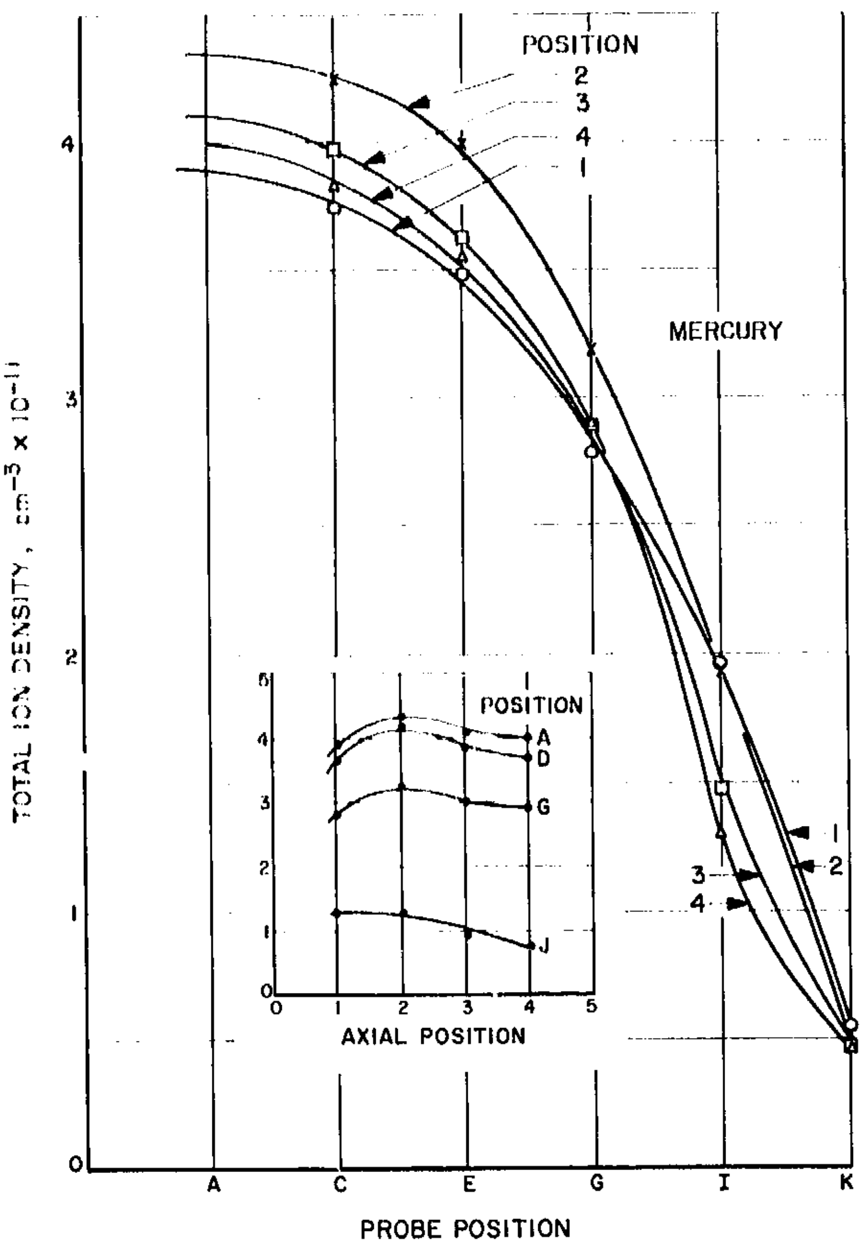


Fig. 31. Typical ion density distribution for mercury

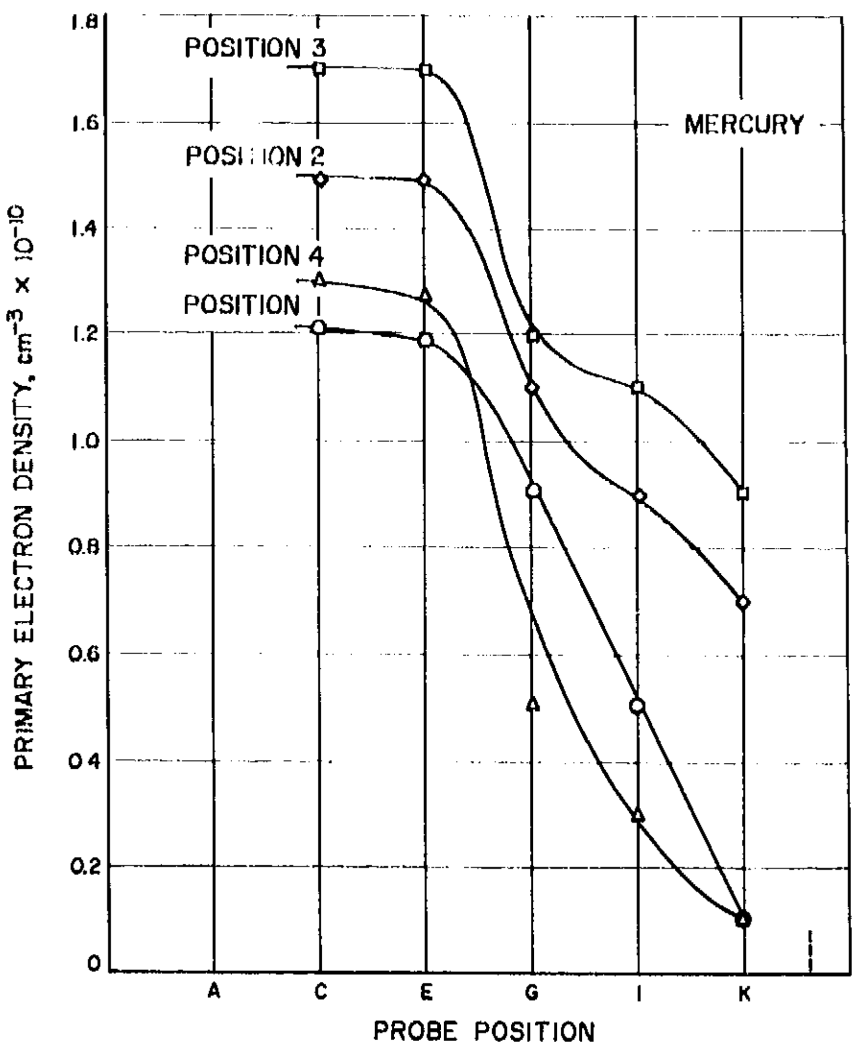


Fig. 32. Typical primary electron density distribution for a mercury plasma

It is also shown in Ref. 38 that the primary electron group is one to two orders of magnitude more effective in producing ionization in mercury than is the Maxwell group. In cesium the same disparity should exist and to a greater extent. Further, the lower ionization cross-section for cesium should result in a higher current requirement to produce the same degree of ionization, which it indeed does, and by approximately the correct ratio to account for the difference in cross-sections (about a factor of two) and the smaller effectiveness of the Maxwells. The important point to be noted, however, is the absence of any sign of a primary group in cesium.

This latter point can be explained by considering the following argument based on collision times. The mean time τ_a between electron-atom collisions is given by

$$\tau_a = \frac{1}{N_a S}\left(\frac{m_e}{2kT_e}\right)^{1/2} \qquad (2)$$

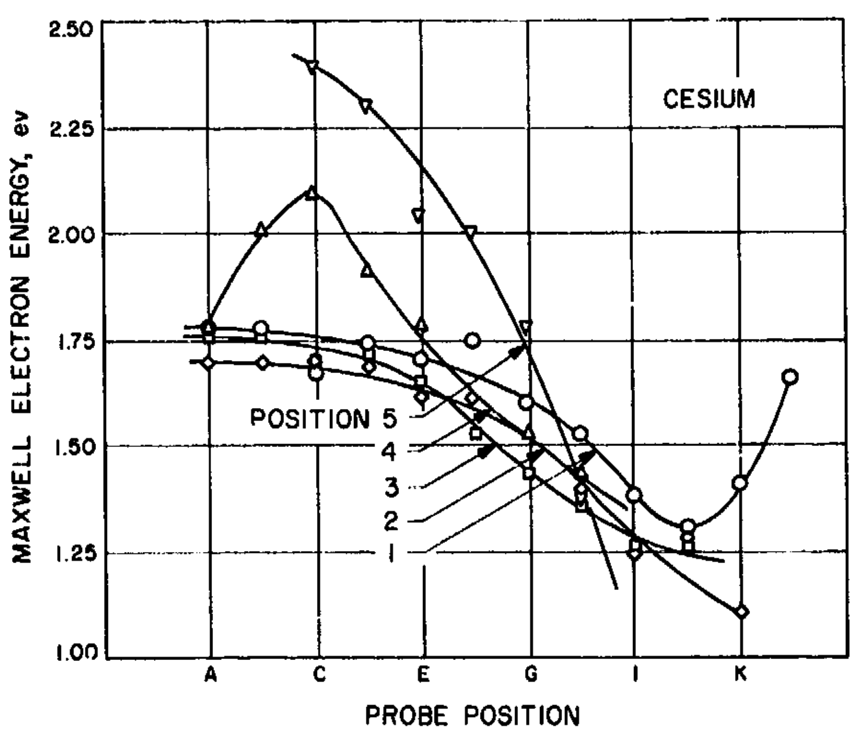


Fig. 33. Typical Maxwell electron energy distribution for cesium

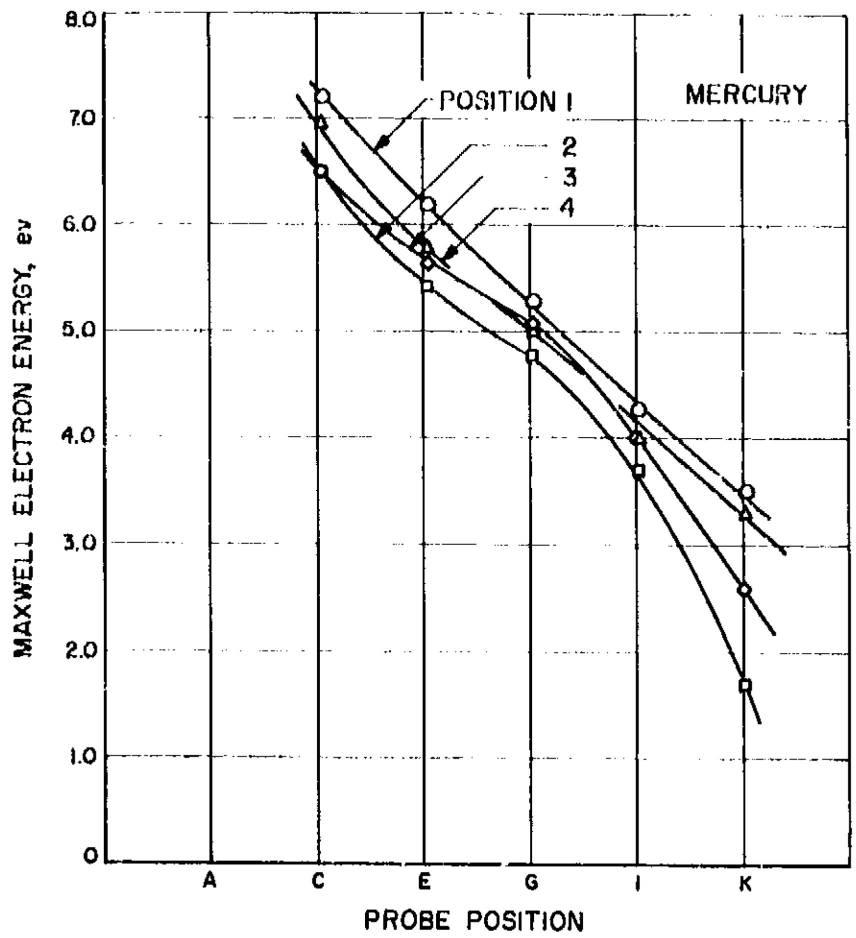


Fig. 34. Typical Maxwell electron energy distribution for mercury

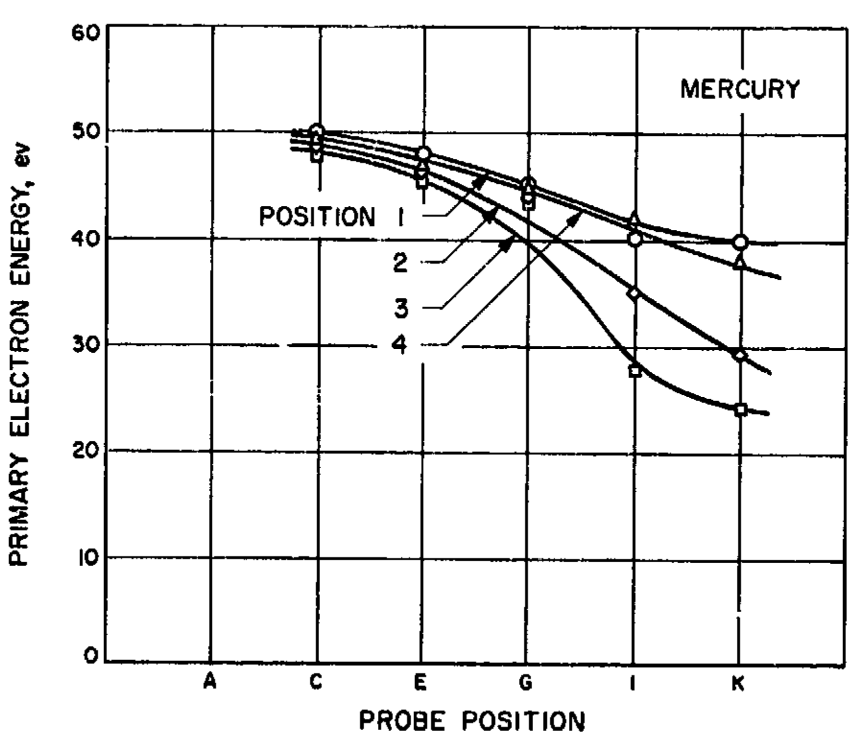


Fig. 35. Typical primary electron energy distribution for a mercury plasma

where S is the electron-atom total collision cross-section, N_0 the neutral particle density, k is the Boltzmann constant, and T_e the electron temperature. The mean electron-electron collision time τ_e is given by Eq. (2).

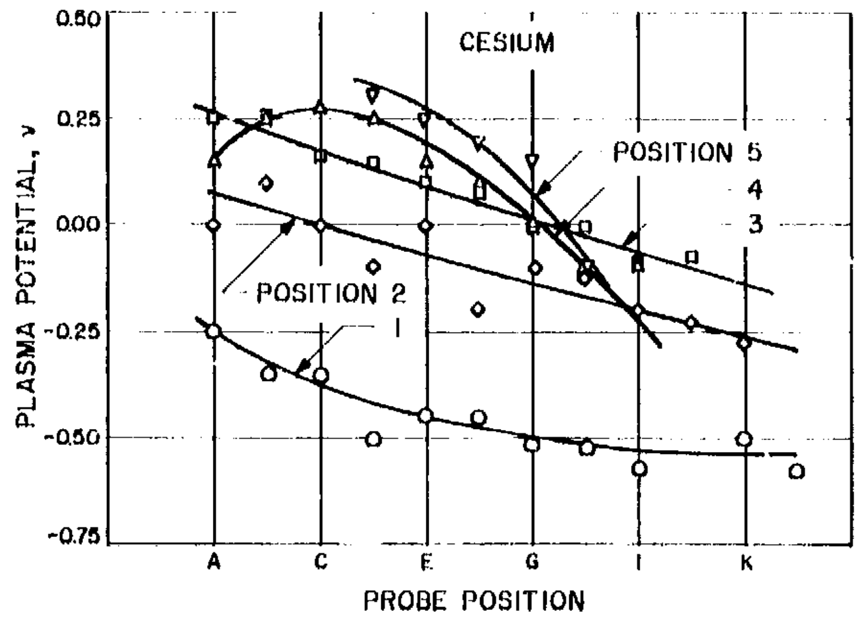


Fig. 36. Typical plasma potential with respect to anode potential for cesium

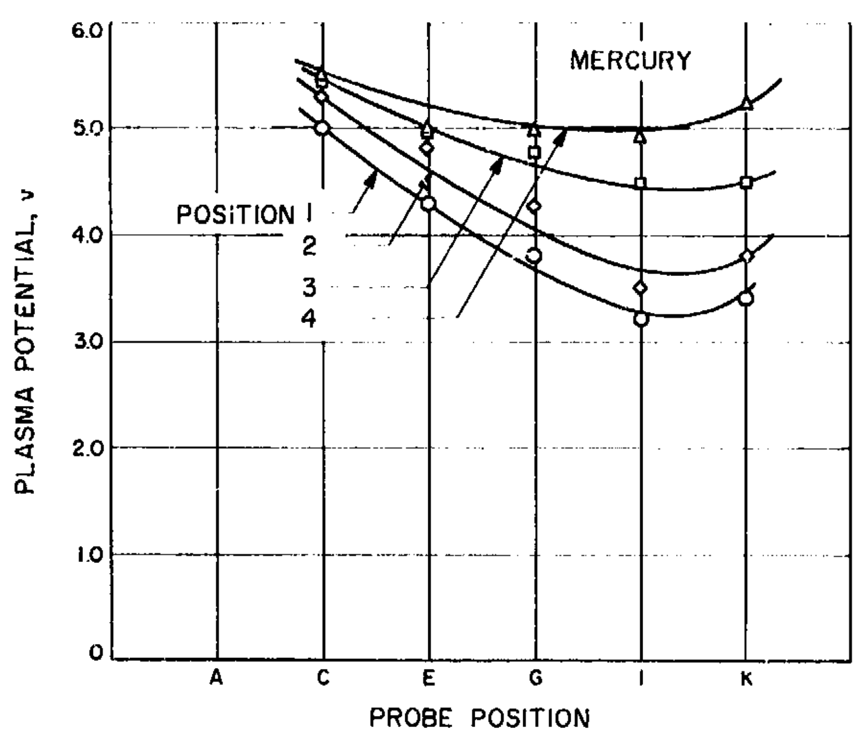


Fig. 37. Typical plasma potential with respect to anode potential for mercury

$$\tau_e = \frac{0.267\, T_e^{3/2}}{N_e \log \Lambda} \tag{3}$$

where N_e is the electron density

$$\Lambda = 12\, (\epsilon_0\, k\, T_e)^{3/2} / e^3\, N_e^{1/2}$$

and ϵ_0 is the permittivity of free space. Since the primary energy loss mechanism for electrons is in electron-electron collisions, the smaller the ratio the more rapidly primaries will give up their energy to other electrons and join the Maxwellian group.

Table 2 lists τ_e and τ_0 for mercury and cesium. It is seen that the ratio differs by two orders of magnitude which would explain the rapid disappearance of the primary beam in cesium and its persistence in mercury.

Table 2. Collision time comparison

	τ_e	τ_0	τ_e/τ_0	T_e	N_e	N_0
Hg	1×10^{5}	1×10^{3}	1×10^{2}	5.0 ev	10^{11}	10^{10}
Cs	8×10^{7}	8×10^{3}	1×10^{1}	1.0 ev	10^{11}	10^{10}

From the foregoing it is apparent that the cause of the ion density distribution in cesium cannot be explained as it can in mercury by the decay of the primary beam across the plasma. Rather, one would expect the greatest portion of the ionization to be taking place near the cathode with electrons and ions diffusing across the chamber. The diffusion coefficient of electrons across a magnetic field is given approximately by

$$D = \frac{C}{B^2} \tag{4}$$

where C is a constant and B is the magnetic field strength (Ref. 39). Experimentally it was found that the cesium plasma operated best when the magnetic field was about 6 to 10 gauss, compared with about 20 gauss for mercury. Thus it would seem reasonable to conclude that diffusion is primarily responsible for the observed distribution in the cesium plasma, while the decay of the primary electron density is the primary mechanism in mercury.

Since the ion density distributions are the quantities of primary interest because of their influence on engine efficiency and lifetime (Ref. 40), it is important to recognize the implications of the foregoing analysis. The mercury and cesium engines in their present configuration, perhaps due to a set of fortuitous circumstances, present strikingly similar results when examined from the outside, as for example, by a beam probe; however, methods to improve the density distributions and enhance engine performance must be tailored to the cause of the distribution. Thus, a more divergent magnetic field, which could carry primary electrons more rapidly across the mercury plasma, might tend to smooth out the distribution in mercury, but would have virtually no effect on the cesium plasma. In the latter case, some means is required to increase the diffusion rate or modify the engine configuration to take better advantage of the higher diffusion rate in cesium. Future activities on this engine will be concentrated in this area.

References

1. Denesse, J. F., and Rocard, Y., *Journal de Physique et le Radium*, Vol. 12, pp. 893–899, 1951.
2. Sen, H. K., *Physical Review*, Vol. 102, p. 5, 1956.
3. Marshall, W., *Proceedings of the Royal Society*, Vol. A233, p. 367, 1955.
4. Jukes, J. O., *Journal of Fluid Mechanics*, Vol. 3, p. 275, 1957.
5. Kolbel, J., Institut fuer Theoretische Physik, University of Hannover, Germany, Technical Report AF 61(052)-161, 1961.
6. Tidman, D. A., *Physical Review*, Vol. 111, p. 1439, 1958.
7. Shafranov, V. D., *Zhurnal Experimentalnoi Teoreticheskoi Fiziki*, SSSR, Vol. 32, p. 1453, 1957.
8. Zel'Dovich, Ia. B., *Zhurnal Experimentalnoi Teoreticheskoi Fiziki*, SSSR, Vol. 32, p. 1126, 1957.

References (Cont'd)

9. Imshennik, V. S., *Zhurnal Experimentalnoi Teoreticheskoi Fiziki*, SSSR, Vol. 42, p. 236, 1962.

10. Prokofev, V. A., *Uchenye Zapiski* (Scientific Reports of the Moscow State University), Vol. 172, Mechanics Section, p. 79, 1954.

11. Grewal, M. S., Institute of Engineering Research, University of California at Berkeley, Technical Report HE-150-198, 1962.

12. Griem, H. R., et al., *Physical Review*, Vol. 116, p. 2, 1959.

13. Petschek, H., and Byron, S., *Annals of Physics*, Vol. 1, p. 270, 1957.

14. Harwell, K. E., "Initial Ionization Rates in Argon, Krypton and Xenon," Thesis, California Institute of Technology, Pasadena, California, 1963.

15. Kerrisk, D. J., and Masek, T. D., "Effects of Plasma Nonuniformity on Grid Erosion in an Electron Bombardment Ion Engine," AIAA Preprint 64-688, presented at the Electric Propulsion Specialists Conference, Philadelphia, Pa., on August 31–September 2, 1964.

16. Speiser, R. C., Sohl, G., and Reid, G. C., "Ion Rocket System Research and Development," Electro-Optical Systems, Inc., Contract Report NAS3-5250, December 20, 1964.

17. Eckhardt, W. O., Hyman, J., Hagen, G., Buckecy, C. R., and Knechtly, R. C., Hughes Research Laboratories, Contract Report NAS3-2511 and NAS3-3564, "Research Investigation of Ion Beam Formation From Electron Bombardment Ion Sources," March 1964.

18. Rapp, D., and Francis, W. E., "Charge Exchange Between Gaseous Ions and Atoms," *Journal of Chemical Physics*, Vol. 37, pp. 2631–2645, 1962.

19. Wehner, G. K., and Rosenberg, D., "Mercury Ion Beam Sputtering of Metals at Energies 4–15 kev," *Journal of Applied Physics*, Vol. 32, pp. 887–890, 1961.

20. Wehner, G. K., "Low-Energy Sputtering Yields in Hg," *Physical Review*, Vol. 112, pp. 1120–1124, 1958.

21. Wehner, G. K., "Sputtering Yields for Normally Incident Hg^{+} Ion Bombardment at Low Ion Energy," *Physical Review*, Vol. 108, pp. 35–46, 1957.

22. Elliott, D. G., Cerini, D. J., and Weinberg, E., *Investigation of Liquid MHD Power Conversion*, Third Biennial Aerospace Power Systems Conference, Philadelphia, Pa., September 1–4, 1964, AIAA Paper No. 64-760.

23. Elliott, D. G., "DC Liquid-Metal Magnetohydrodynamic Power Generation," *Proceedings of Sixth Symposium on the Engineering Aspects of Magnetohydrodynamics*, University of Pittsburgh, Pittsburgh, Pa., April 21–22, 1965.

24. Elliott, D., Cerini, D., Hays, L., O'Connor, D., and Weinberg, E., "Liquid MHD Power Conversion," *SPS 37-28*, Vol. IV, pp. 86–92, Jet Propulsion Laboratory, Pasadena, California, August 31, 1964.

25. Elliott, D., Cerini, D., and Eddington, R., "Liquid MHD Power Conversion," *SPS 37-29*, Vol. IV, pp. 106–115, Jet Propulsion Laboratory, Pasadena, California, October 31, 1964.

References (Cont'd)

26. Elliott, D., Cerini, D., Hays, L., and Weinberg, E., "Liquid MHD Power Conversion," *SPS 37-30*, Vol. IV, pp. 116–119, Jet Propulsion Laboratory, Pasadena, California, December 31, 1964.

27. Elliott, D., Cerini, D., Hays, L., and Weinberg, E., "Liquid MHD Power Conversion," *SPS 37-31*, Vol. IV, pp. 186–190, Jet Propulsion Laboratory, Pasadena, California, February 28, 1965.

28. Goldsmith, A., Waterman, T., and Hirschhorn, H., *Handbook of Thermophysical Properties of Solid Materials*, The Macmillan Co., New York, N.Y., 1960.

29. Lewins, J., "Time-Dependent Variational Principles for Nonconservative Systems," *Nuclear Science and Engineering*, Vol. 20, pp. 517–520, 1964.

30. Lewins, J., "A Variational Principle for Nonlinear Systems," *Nuclear Science and Engineering*, Vol. 12, pp. 10–14, 1962.

31. Becker, M., and Fenech, H., "Least-Squares Variational Methods," *Transactions of the American Nuclear Society*, Vol. 7, No. 1, p. 11, 1964.

32. Pomraning, G. C., and Clark, M., Jr., "The Variational Method Applied to the Monoenergetic Boltzmann Equation," Parts I and II, *Nuclear Science and Engineering*, Vol. 16, pp. 147–164, 1963.

33. Selengut, D. S., "Variational Analysis of a Non-Linear Kinetics Problem," *Transactions of American Nuclear Society*, Vol. 7, No. 2, pp. 256–257, 1964.

34. Bellman, R., "Dynamic Programming," Princeton University, Princeton, N. J., 1957.

35. Bellman, R., and Kalaba, R., "Dynamic Programming, Invariant Imbedding and Quasi-linearization: Comparisons and Interconnections in Computing Methods and Optimization Problems," A. V. Balakrishnan and L. W. Neustadt, Editors Academic Press, N.Y., 1964.

36. Wing, G. M., "An Introduction to Transport Theory," John Wiley and Sons, N.Y., 1962.

37. Kaplan, S., Marlowe, O. J., and Bewick, J., "Application of Synthesis Techniques to Problems Involving Time Dependence," *Nuclear Science and Engineering*, Vol. 18, pp. 163–176, 1964.

38. Strickfaden, W. B., and Geiler, K. L., "Probe Measurements in an Operating Electron Bombardment Engine," *AIAA Journal*, Vol. 1, p. 1815, 1963.

39. Spitzer, L., *Physics of Fully Ionized Gases*, Interscience Publishers, New York, N.Y. 1962.

40. Kerrisk, D. J., and Masek, T. D., "Effects of Plasma Non-Uniformity on Grid Erosion in an Electron Bombardment Ion Engine," AIAA Preprint 64-688, 1963.

XV. Liquid Propulsion

A. Advanced Liquid Propulsion Systems

R. N. Porter, D. L. Bond, H. B. Stanford, W. F. MacGlashan, and O. F. Keller

1. *Introduction,* R. N. Porter

The Advanced Liquid Propulsion Systems (ALPS) program is investigating selected problems generated by spacecraft operational requirements for propulsion systems capable of high inherent reliability, long-term storage in space, multiple start in free fall (zero-gravity), and engine throttling. The solutions proposed to satisfy these requirements have been incorporated into the ALPS system.

Periodic reports in the *SPS*'s (starting with *SPS 37-8*) describe the progress of work on the various parts of the ALPS system. Recent accomplishments are outlined below. These include activation of a new six-component thrust-measuring system, additional crease tests of potential bladder materials, bipropellant expulsion tests using nested convoluted diaphragms, development of a liquid regulator, and a tank and expulsion cell assembly.

2. *System Development,* R. N. Porter

The basic ALPS system concept is shown diagrammatically in Fig. 1. It features a variable-thrust rocket engine, fed by two positive-expulsion devices contained within a single vessel pressurized by gas generated through the decomposition of a monopropellant. This

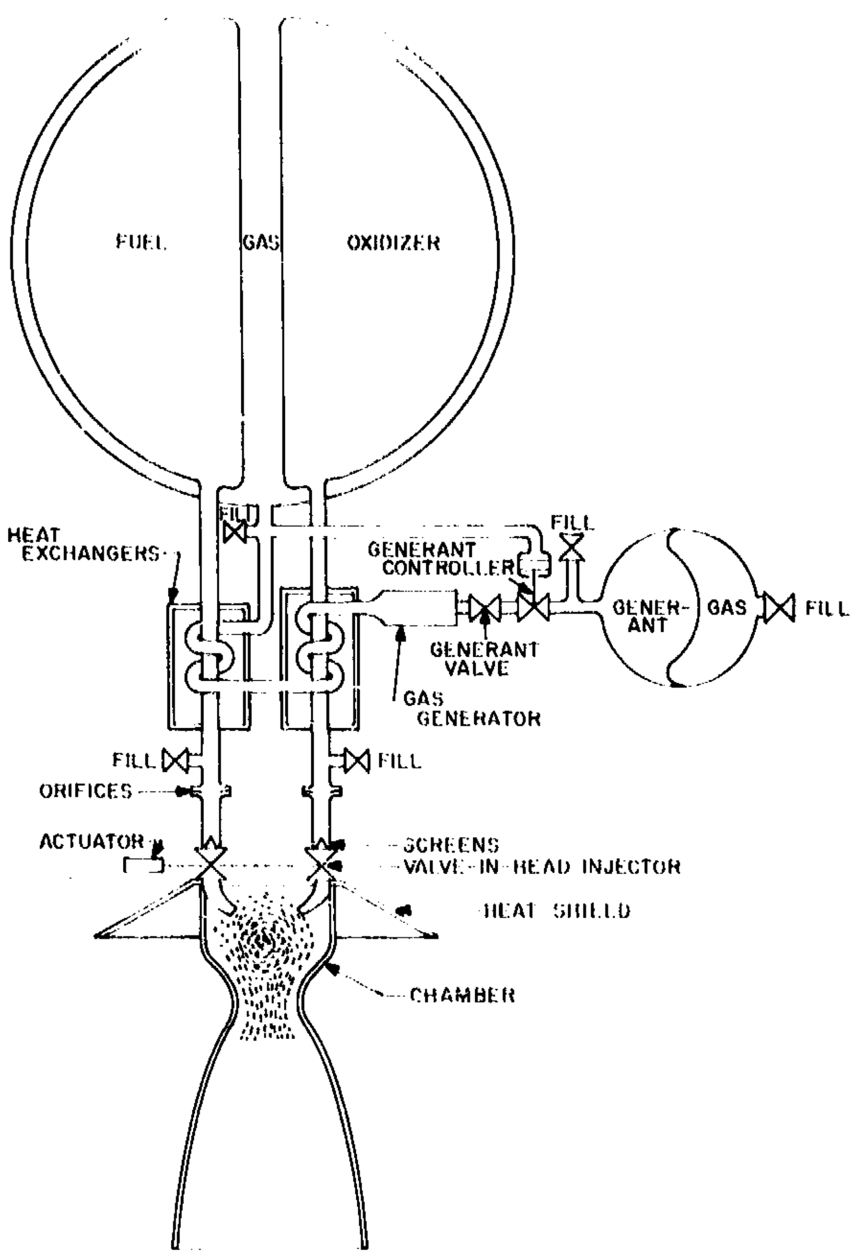


Fig. 1. Schematic diagram of the ALPS system

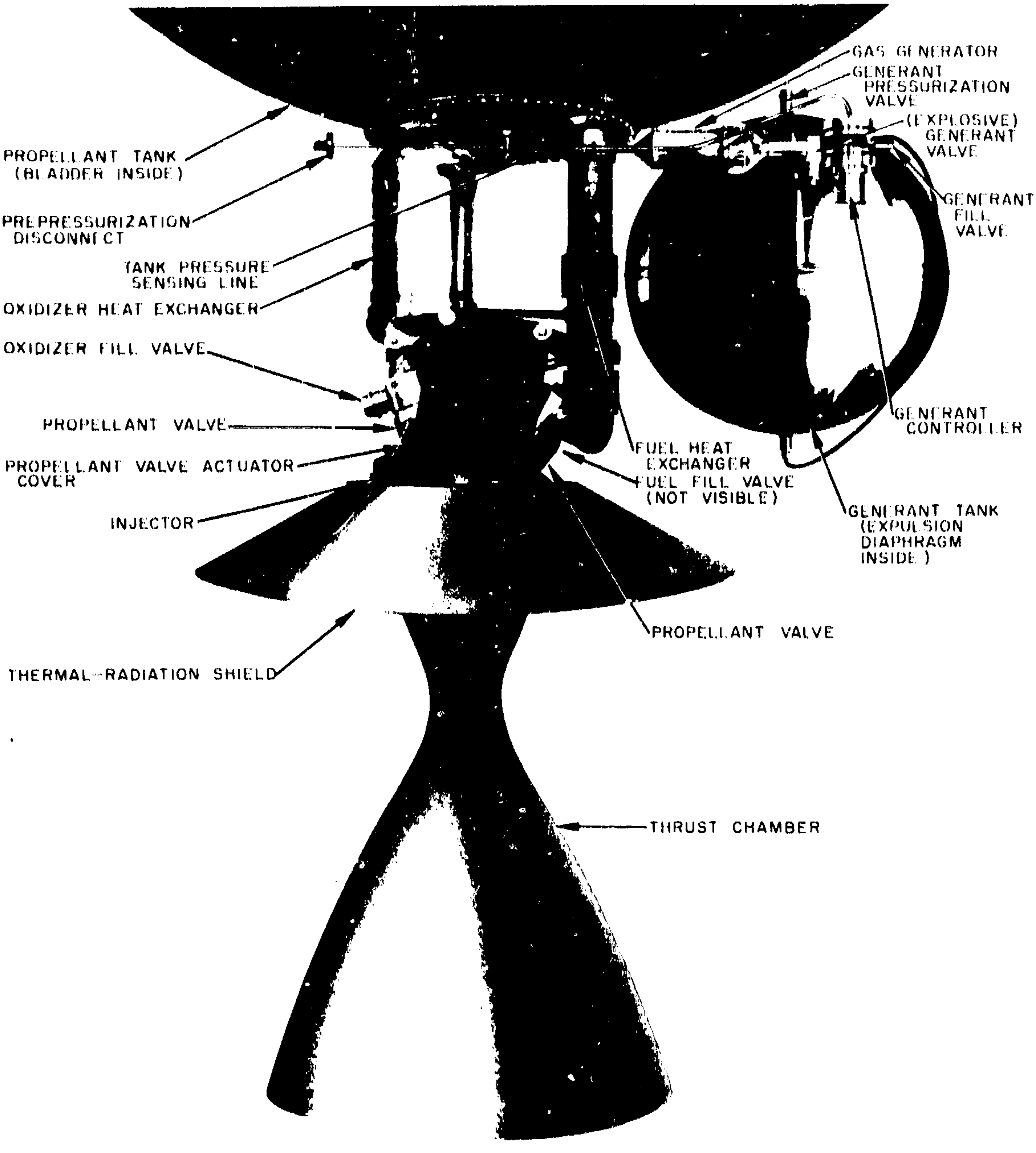


Fig. 2. A possible arrangement of the ALPS system

system was devised in an attempt to maximize reliability, versatility, controllability, and storability, and to minimize interactions with the rest of the spacecraft.

At this writing, no complete ALPS system has been assembled. However, a considerable amount of analytical work has been done, using analog and digital computer programs to predict the center-of-gravity excursions, the variations in operational parameters for various conditions (acceleration, temperature, and thrust) and the transient behavior. The results of these analyses indicate that the system would operate within very narrow limits of mixture-ratio, chamber-pressure, and center-of-gravity excursions (assuming no propellant sloshing) and would be capable of rapid thrust transients.

Several layouts of the system have been made. One version, assembled from existing ALPS component parts, is shown as a composite photograph in Fig. 2.

Since a complete system has not been available, tests have been limited to subsystems. Feasibility of the pressurization-circuit concept was successfully demonstrated in early 1961. A throttleable gas generator fed by a mockup of the ALPS generant feed circuit was used to pressurize a heavyweight test tank from which water was allowed to flow at varying rates to simulate the modulated flow of propellants to an engine. Similar tests are being made of the latest ALPS hardware.

3. *Thrust Chamber Testing*, D. L. Bond

A new multicomponent thrust-measuring system has been in service at the Laboratory's Edwards Test Station since January 1965. Multicomponent thrust measurement is particularly useful when testing ablative, or other nozzles which change geometry during the firing period, because uneven erosion can be detected by changes in the thrust vector.

The new system consists of a suspended thrust mount which is connected by flexures and load cells to a firmly anchored rigid frame. Six Schaevitz-Bytrex solid-state load cells are used to measure six components of the thrust generated by the rocket engine. Nominal axial thrusts of up to 2,000 lb can be accommodated. These measurements can be made on chambers exhausting directly to the atmosphere, as in Fig. 3, or to low pressures by moving the system into the Altitude Simulation Facility.

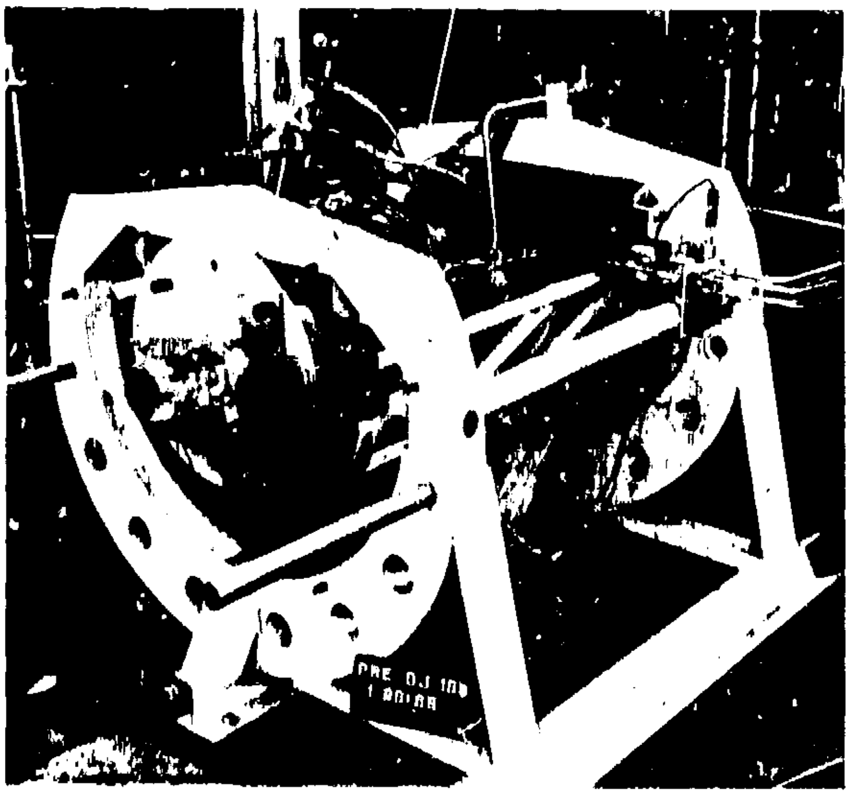


Fig. 3. 100-lb thrust ablative chamber mounted on the six-component thrust-measuring system

Based on data from the calibrations and firing tests made to date, the performance of the system appears to be reasonably good and capable of improvement. Analysis of the data indicates axial measurements yield values for nozzle thrust coefficient C_f which are consistent to within ±0.2%. Reading accuracy of the side loads is ±0.3% of full scale.

Two problems are apparent. One arises from the interactions between the side and axial loads. Applying a pure axial load generates a side force as large as 1% of the applied load, while a pure side load produces an axial force amounting to as much as 3% of the side load. Correction of this deficiency will probably involve a change in the method of mounting the load cells. A more serious problem is the "ringing" which obscures the measurements of thrust transients. When fitted with a 100-lb thrust rocket engine and the appropriate load cells, the system's suspended mass is about 240 lb and the natural frequency is 40 cps. As a temporary expedient, electrical filtering of the signals has been used to remove the spurious output caused by the mechanical resonance of the system. Current thrust-chamber test firings are being conducted with this system to determine changes in thrust vector over periods of hundreds of seconds, therefore the ringing does not interfere with securing the desired information; however, future tests of thrust decay transients and throttling will require elimination of the ringing to secure accurate data.

4. Expulsion Devices Development, *H. B. Stanford*

The basic design of the Advanced Liquid Propulsion System specifies that the hypergolic propellants (N_2O_4 and N_2H_4) be stored in and expelled from a single spherical tank. To meet these requirements and those imposed by a space environment, the propellants must be separated and controlled by an expulsion device. Compatible, impermeable, and durable bladders are considered to best meet these requirements and are the first choice for use in the ALPS system. Finding suitable materials for these bladders has proved difficult. One important phase of this search is the crease testing of potential materials; recent crease test data are discussed immediately below. Due to difficulty in obtaining a completely satisfactory bladder, a backup development was considered necessary. This second choice is the convoluted metal diaphragm. Later paragraphs in this report describe a test involving the simultaneous expulsion of both fuel and oxidizer from a single tank with diaphragms of a nested design.

a. Crease-testing bladder materials. In the Advanced Liquid Propulsion System bladder development program, material selection and testing is considered to be of prime importance. One phase of this study has been an investigation of the relative resistance of materials to the forces which cause tears and pinholes as folds and creases are induced during expulsion cycling. To perform tests of this nature with actual bladders is both time consuming and costly. A search for a commercial tester that would accurately simulate creases, as formed during the collapse of an expulsion bladder, failed to disclose one. Therefore, a device was designed at JPL which generates creases in material samples very similar to those found in an expulsion bladder. The machine appears to do this in a reproducible manner from sample to sample, thereby giving good relative information on the failure resistance of the samples tested. A description of the machine and the theory of its operation were published in *SPS 37-28* and *37-24*, Vol. IV.

In the most recent tests of potential bladder materials, the samples were folded over a pair of flat blades inclined at 30 deg so that a tight fold was made in the samples at the vertex of the angle. The ends of the samples were attached to cables with a preset tension of 5 lb. The eccentric drive pulled the samples back and forth with a stroke of approximately 3 in. at a rate of 30 cycles/min. The sample was determined to have failed when tears, delamination, etc., were observed or when a dye penetrant, which was periodically applied, disclosed a leak. The results of these tests are shown in Table 1. Many of the samples listed were supplied by the Dilectrix Corporation of Farmingdale, Long Island, New York. Some of these materials, particularly the "codispersions" and "redundant" films, show remarkable flex life. On the average, these materials are ten to twenty times as resistant to failure from creasing as the common sprayed-and-sintered TFE-FEP bladder material, TFE skived film, etc. Dilectrix describes their codispersion material as TFE-FEP Teflon sprayed and sintered together into a uniform film, and the redundant films as discrete layers of TFE Teflon, built up to about 0.003-in. in thickness, interspersed with 0.0002- to 0.0003-in.-thick layers of TFE mixed with some other material such as "Aqua-DAG" (a finely dispersed graphite) or aluminum flakes. The manufacturer says this film is adaptable to a diversity of structures, since any reasonable number of layers, layer thicknesses, and interspersing layers with additives, can be assembled. FEP Teflon can also be added to reduce permeation, although the samples reported herein did not contain FEP.

It is interesting that the Dilectrix Teflon-aluminum foil laminate sample which is assembled over a sprayed redundant film base endured in excess of 25 crease cycles before failure, which is better than the best Teflon cloth-aluminum foil laminates previously reported (*SPS 37-30*, Vol. IV). Failure in this material was detected by visual observation of fractures in the aluminum foil; the pinhole test with dye penetrant indicated that no leak paths had been formed through the base film. It should be noted, however, that this was a single test and has yet to be repeated.

All of the other materials tested, including those with gold Afflair and aluminum flake, were produced as part of a very limited study being conducted at JPL with the goal of reducing permeation to N_2O_4. These samples were made with FEP Teflon only, and were not expected to exhibit high crease resistance. All of these samples failed very quickly.

b. Convoluted diaphragms. A continuing development effort on metal diaphragms has been maintained at JPL for the past few years, and some progress has been made, particularly in simultaneous expulsion of hypergolic propellants from a single tank. Initially this was done with identical diaphragms which were placed back to back in an 18-in.-D spherical tank with a perforated metal plate between them. In this manner simultaneous expulsion with the propellants N_2O_4 and N_2H_4 were accomplished twice without mishap. A brief description of these tests appeared in *SPS 37-27*, Vol. IV.

Table 1. Results of crease tests of potential bladder materials

Source and material	Thickness	No. of cycles to failure	Type of failure
Dilectrix, Teflon and aluminum foil laminate	Al 0.0005 in.[a] FEP 0.003 in.[a] TFE 0.012 in.[a] Total 0.0155 in.[b]	25>50	Fine line crack in aluminum
Dilectrix, TFE-FEP codispersion, R&D No. 40	0.010 in.	1400>1497	Tear
		1700>1800	Tear
		1900>2000	1 pinhole
Dilectrix, redundant film structure, R&D No. 41	3-0.003-in.[a] thick layers of TFE and 2-0.0002-in.[a] thick layers of TFE mixed with Aqua-DAG Total 0.0095[b]	3000>3100	Several pinholes
		2900>3000	1 pinhole
		1900>3000	1 pinhole
Dilectrix, redundant film structure, R&D No. 42	2-0.003-in.[a] thick layers of TFE and one 0.0002-in.[a] thick layer of TFE mixed with Aqua-DAG Total 0.0065 in.[b]	1800>1900	1 pinhole
Dilectrix, redundant film structure, R&D No. 42	2-0.003-in.[a] thick layers of TFE and one 0.0002-in.[a] thick layer of TFE mixed with Aqua-DAG Total 0.0065 in.[b]	3900>4000	6 very fine pinholes
		3300>3400	3 small pinholes
Dilectrix, redundant film structure, R&D No. 27	3-0.003-in.[a] thick layers of TFE and 2-0.0003-in.[a] thick layers of TFE mixed with aluminum flakes Total 0.010 in.[b]	1900>2000	Tears in two places
		1800>1900	1 pinhole
		2300>2400	6 pinholes

Source and material	Thickness	No. of cycles to failure	Type of failure
Dilectrix, TFE-FEP sprayed and sintered bladder material, cut from No. 1878 tapered bladder	Tapered 0.010 to 0.023 in.	50>75	1 pinhole
		225>250	1 pinhole
Commercial, molded and skived TFE	0.010 in.	25>30	Several pinholes
JPL, FEP mixed with mica, sprayed and sintered, JPL No. 30	0.0065 in.	10>20	3 pinholes
JPL, FEP mixed with gold Afflair, sprayed and sintered, JPL No. 28	0.006 in.	10>20	2 pinholes
JPL, FEP mixed with gold Afflair, sprayed and sintered, JPL No. 29	0.007 in.	30>40	3 pinholes
JPL, FEP mixed with aluminum flake, sprayed and sintered, JPL No. 31	0.007 in.	10>20	3 pinholes
JPL, FEP mixed with aluminum flake, sprayed and sintered, JPL No. 32	0.0045 in.	20>30	2 pinholes

[a]Layer dimensions from vendor description
[b]Total figures, actual measurement by micrometer

When using two like diaphragms, 7 to 9% of the tank volume is contained between them which therefore is unavailable for propellant storage. Nesting the diaphragms (Fig. 4), reduces the unusable space to approximately 3% of tank volume. 18-in. diaphragms were designed in order to procure items of this configuration for testing in existing equipment. A major difficulty was encountered in determining the proper convoluted shape, having large radii at the crests and valleys (inherent to the spinning process) that would result in the formation of a nearly perfect hemisphere upon expulsion. Feeling that oversize hemispheres were more desirable than undersized ones, an extra ⅜ in. was added to the developed length of the contour. Phoenix Products of Milwaukee, Wisconsin, fabricated the diaphragms, using a spinning process and temporary wooden tooling. Certain limitations seem to be inherent in the spinning process, one being the necessity for large radii at the convolution crests as mentioned above, and another being a variation in thickness which, for these diaphragms, ranged from approximately 0.060 in. in the center to 0.025 in. at the outer convolution. This thickness variation can be reduced by grinding, but these diaphragms were received in the "as-spun" condition at JPL's request, except for sanding to remove major surface blemishes.

Expulsion tests were run with single diaphragms, first with water in a plexiglass tank (pressure limitation, 30 psi), and then with each propellant in a stainless steel tank. The diaphragms were a tight fit in both tanks, and it was noted that after expulsion the fuel diaphragm (deep convolution) was wrinkled, indicating that the built-in oversize condition was present. No wrinkling was noted in the oxidizer diaphragm.

Upon completion of the single-fluid expulsion tests a nested pair of diaphragms were installed in the stainless steel tank for a bipropellant expulsion test. The tank was mounted in the slosh test rig at Edwards Test Station (*SPS 37-27*, Vol. IV) and the fuel side filled with N_2H_4. After a 24-hr period the oxidizer side of the tank and the space between the diaphragms were carefully checked for fuel vapor which would have indicated a leak. When no fuel vapor was found, the oxidizer side was remotely filled with N_2O_4, the slosh rig adjusted to an amplitude of ± 1½-in. lateral motion at 128 cycles per minute, and the expulsion test immediately conducted at a slosh acceleration of 1 g.

Nothing unusual was noted until the very end of the run, at which time tank pressure decayed and the oxidizer flowmeter, instead of indicating zero flow, ran wild.

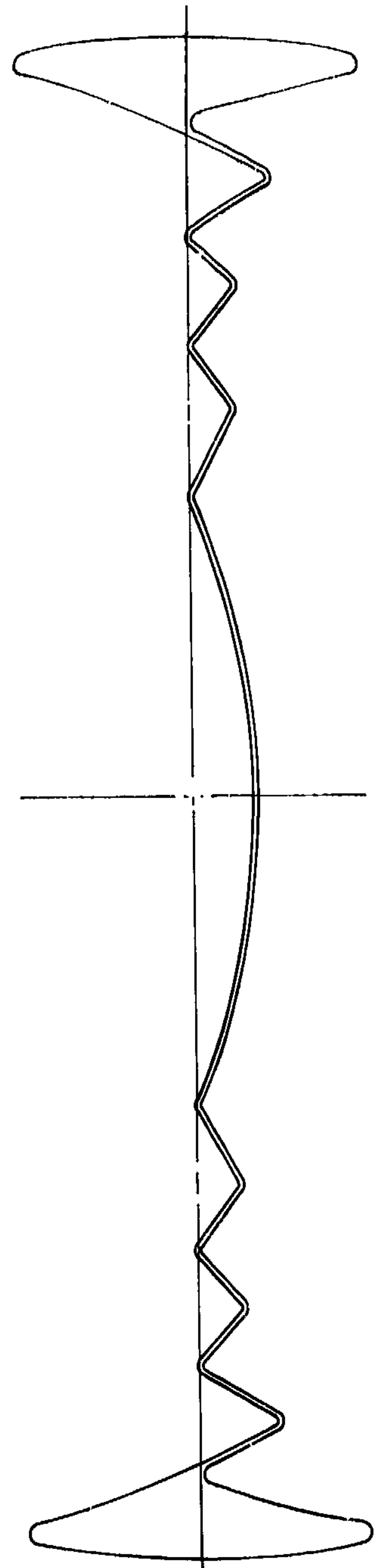


Fig. 4. 18-in. D nested pair of convoluted diaphragms used for bipropellant expulsion

Upon disassembly it was found that the fuel diaphragm was intact but wrinkled, as had been noted in previous expulsion tests. The oxidizer diaphragm, however, was split for a distance of about 6 in. along the crest of the outer convolution. Measurement of the developed length of the contour from flange to flange of the distended diaphragm was ⅕ in. greater than that of similar diaphragms in the as-spun condition. As near as could be determined, these undistended diaphragms were nominal for 18-in. D and did not have the planned for ⅜ in. oversize on the flange-to-flange dimension. Close examination of the fractured area revealed thinning to approximately 0.015-in. thickness due to high tool pressure, deep tool marks (these diaphragms were in the as-spun condition) and very obvious necking down of the metal on either edge of the break. Hardness and bend tests indicated the metal to be only slightly harder than fully annealed 1100 aluminum, and microscopic examination showed the material to be clean.

From these facts it can be conjectured that the diaphragm split due to stretching of the metal rather than from bending. The stretching was permitted by the fact that the diaphragm was undersize (both from the design dimension and for the tank in which it was tested), the tank having been built to a plus tolerance, also was not perfectly spherical. The fracture occurred at a point on the diaphragm that had been weakened by thinning. This condition was aggravated by deep tool marks. Since no indication of failure was evident until the end of the run, it is assumed that the diaphragm split at this time; although this fact cannot be proven.

It can, therefore, be concluded that this failure was due to dimensionally mismatched components and to a metal thinning condition caused in manufacture. Both of these conditions are readily correctable, and adequate inspection of integral test items would preclude reoccurrence of this condition. Despite the difficulties encountered, the spinning process is considered a satisfactory method of diaphragm manufacture within its stated limitations. From all other aspects the simultaneous expulsion of the propellants N_2O_4 and N_2H_4 from spun nested diaphragms was a success.

5. Generant Controller Development

W. F. MacGlashan

This is the first report on the progress of the generant controller which has been under development at JPL for several years.

a. Function and requirements. In the ALPS system the propellant tank is pressurized with gases generated by the decomposition of hydrazine in a gas generator. The function of the generant controller is to meter the flow of hydrazine (or generant) to the gas generator, so that constant pressure is maintained in the propellant tank during all firing periods, regardless of whatever variations may occur in the rates at which the propellants flow out of the tank.

The current design is sized for a nominal generant flow rate of 0.1384 lb/sec at a pressure drop across the controller of 75 psi, or less. Since engine throttling and other factors may require substantially greater or lesser flows, the controller must be able to operate precisely and stably over a flow rate range of 20 to 1. Inlet pressure to the generant controller varies from a maximum of 1500 psi to a minimum of 500 psi, because it operates on the "blowdown" principle (see Sec. 6, *Generant Tank and Cell*, for a definition of this principle). Delivered outlet pressure from the controller is not a significant design parameter in the ALPS application, since the controller is insensitive to it. Rather, propellant tank pressure, which the controller senses and to which it responds, will be set somewhere between 200 and 300 psi; whatever preset pressure is chosen, however, should be maintained by the controller within very narrow limits. External leakage of the fluids at these pressures must be kept below 10 cm^3/yr (stp) of nitrogen from the sensing mechanism and below 1 cm^3/yr of hydrazine from the liquid-containing part of the controller.

The generant controller must perform reliably after storage periods of up to one year in a spacecraft environment. This means that the exterior of the controller has to be immune to high vacuum and a dose of radiation while the interior must be compatible with hydrazine (and water for testing). Operating temperature of the ALPS system is 70 ±30°F, so no severe temperature cycling problems will be faced, unless the controller is used in a system which must be heat sterilized. If sterilization is necessary, three cycles in which the temperature is raised from ambient to 300°F, held for 24 hr at that level, and lowered back to ambient, must be survived without damage or change in performance. It appeared that an all-metal design would be most likely to withstand these environmental conditions.

b. Design and theory of operation. The ALPS controller is essentially a remote-sensing, single-stage, spring-loaded regulator. Three basic parts comprise the generant controller: a valve through which the generant flows, a pressure sensing piston, and a spring assembly. Flow through the controller is varied by changing the annular gap between a moveable ball and a fixed circular seat in the valve. The ball is positioned by a push rod, which in turn is acted on through adjacent parts by a flat plate. This plate, against which the spring assembly pushes, is tied by three rods to the sensing piston to form a rigid yoke. Regulation of the flow is accomplished because the preset reference force exerted by the spring assembly and the force developed by the sensed pressure on the piston are acting in opposition to each other through the yoke, so that the net difference between the two forces tends to push the ball to one side or the other of the equilibrium position. In other words, when the sensed pressure is low, the pressure force is less than the spring force, and the valve is opened. When the pressure force is high, it overcomes the spring force, and the valve is closed.

The major design problem in such a device is eliminating all other unbalanced forces which might enter into the force balance. For example, in order to be able to transmit the motion from the yoke to the ball, a moveable diaphragm is used to seal the end of the liquid passage, and the liquid pressure generates a force on the diaphragm. Also, another force is created on the ball by the difference between the upstream and downstream pressures. These forces are balanced out by making the projected area of the diaphragm and the seat exactly the same and by sealing the other end of the liquid passage with a diaphragm of the same area upon which the liquid exerts a force. By transmitting these several forces through members held rigidly in contact, they counterbalance one another. Obviously, it is also necessary to minimize friction forces to obtain accurate regulation. The means taken to accomplish this are mentioned below in the paragraphs which describe the individual elements of the controller.

*c. **Ball and seat.*** The ceramic ball is made to very close tolerances. Sphericity is controlled to within 25 μin. Surface finish is 1 to 2 μin. RMS. These dimensions, which should be altered but little in service, because of the hardness and corrosion resistance of the ceramic ball, contribute to the accuracy with which the ball fits the seat. The metal seat, which is a land approximately 0.003-in. wide, is carefully lapped to a very fine finish.

*d. **Diaphragms.*** To provide the flexible hermetic seals which are needed to contain the fluids in the controller, three metal diaphragms are used. Sufficient flexibility, over the very limited stroke required, is obtained by using material which is very thin. However, these thin diaphragms cannot sustain the loads imposed by the pressure forces where the diaphragms bridge the annular gaps without some permanent deformation unless suitable support is provided. Metal backup rings are used to support the diaphragms across the gaps. These rings are described in the next section.

Originally the diaphragms were made from 0.003-in.-thick Type 1100-1 aluminum sheet. Soft aluminum was used for three reasons: (1) soft aluminum eliminates the need for separate gaskets; (2) pure aluminum is compatible with propellants and with water; (3) soft aluminum is fairly resistant to damage from flexure beyond the yield point. The use of aluminum for diaphragms was abandoned because pinholes developed during water tests. This was attributed to galvanic action originating where foreign particles were embedded in the diaphragm surface.

Presently the diaphragms are made from 0.001-in.-thick Type 302 stainless steel shim stock. Because of the greater strength of this hard-rolled material, the thickness was reduced from that previously used. Soft aluminum gaskets, cut from 0.003-in. aluminum sheet stock, are used to seal the edges of the diaphragms. Steel is further down the electromotive list than aluminum, and this may explain why no diaphragm corrosion problem has been detected. In the present design the diaphragms are sealed at the edges by compression between the structural parts. Controllers built in the future will probably have diaphragms sealed by electron beam welding to the adjacent parts.

Diaphragm life has been greatly extended by placing a 0.001-in.-thick stainless steel shim stock backup disc between the backup ring and the diaphragm. The backup disc is free floating and is slightly larger in diameter than the OD of the backup ring.

*e. **Backup rings.*** The backup rings serve the dual functions of radially positioning central moving parts and supporting the diaphragms. These rings are supported on both sides of the annular gaps by narrow bands recessed precisely enough to position the faces of the rings so that they are exactly flush with the surfaces which clamp the edges of the diaphragms.

The large ring which supports the sensing piston diaphragm is radially slit into a number of segments which are left joined together by a very shallow band on the internal diameter. These slits enable the ring to assume a conical shape, as the piston is moved, without offering much resistance. The smaller rings must be separated into completely unattached segments to obtain sufficient flexibility. In the radial cross-section, the rings are contoured with the intent of obtaining uniform stress distribution. At the inner and outer edges, where they are simply supported, the section is thinned to an optimum thickness for the shear loads which are applied. The center section is thickened to resist the bending loads.

Experience has shown that these backup rings are performing as expected. For example, a backup ring made of Type 7075-T6 aluminum seated in an annular gap test fixture of Type 6061-T6 aluminum was used to support a 0.003-in.-thick diaphragm of Type 1100-0 aluminum in a proof-test to a pressure of 5400 psi. There was no damage to any of the parts.

f. Spring assembly. An optimum assembly for providing the preset reference force within the generant controller should be light in weight, have low hysteresis, and have an adjustable spring constant (slope of force versus deflection) so that the controller's regulation characteristic can be matched to specific requirements. Design analysis early in the program showed that conventional leaf, coil, or torsion bar springs would not satisfy all of these requirements simultaneously. It appeared that the type of spring most likely to solve these problems was a stack of Belleville springs.

In the generant controller, the Belleville springs are used in the range of deflections where the slope of the load/deflection curve is negative (i.e., where the load decreases as the deflection increases). Belleville springs are common enough at this time to be commercially procurable. It has been found necessary, however, to rework them to achieve the exact shape and concentricity required for this application. Even after this rework, the load/deflection curve of each spring is slightly different, even though the springs appear to be identical. Each spring is individually calibrated so that when a spring assembly is being put together, those which have exactly the needed characteristics can be selected.

The first spring assembly made for the generant controller consisted of a stack of Belleville springs which was centered on the OD by a sleeve and on the ID by a shaft. The edges of the springs were separated by rings, rolled from round music wire, in order to reduce the friction. It was found that with this arrangement hysteresis was excessive. Next the type of assembly shown in Fig. 5 was

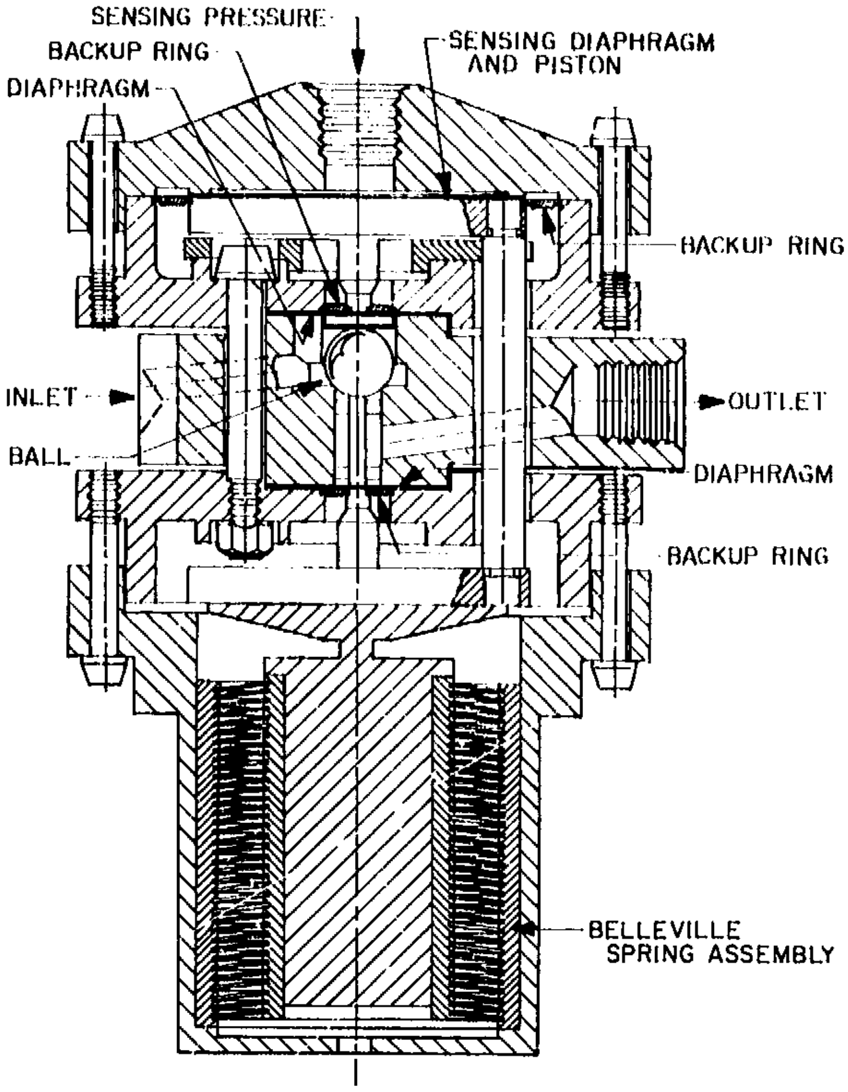


Fig. 5. Cross-section of ALPS generator controller

built. In this design the edges are supported by shallow grooves in two split sleeves. These sleeves are held together by a shaft on the inside and an outer housing. Measurements showed the hysteresis of this assembly to be at an acceptable level.

A further reduction in the hysteresis will be made in a still newer model of the spring package which is now being designed. This latest concept involves supporting each spring edge at only four points. Fig. 6 shows the hysteresis reduction achieved in a single spring with the four-point support as compared to the continuous edge support used in the current design.

g. Performance. Only limited testing of the complete controller has been accomplished to date. Bench tests are of questionable validity because of the remote-sensing nature of the controller's normal operation. One actual firing test of a monopropellant rocket system based on the blowdown feed system concept and using the controller has been made as part of another program. For this use, it was necessary to install the controller so that

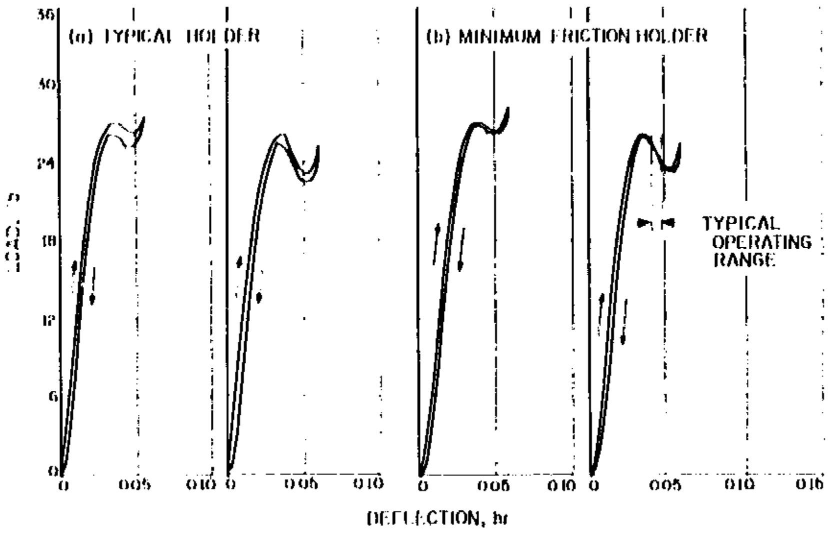


Fig. 6. Load versus deflection curves from two types of Belleville spring supports illustrating the hysteresis in (a) typical holder (b) minimum friction holder

it would deliver a constant hydrazine flow rate at a constant pressure; to do this, the outlet pressure from the controller was fed back to the sensing piston. During the steady-state portion of the 190-sec firing, the generant controller's outlet pressure was a constant 307 psi ±½, while the inlet pressure decayed from approximately 1300 to 600 psi.

A substantial amount of performance data is expected to be gathered during a series of ALPS pressurization circuit tests which was ready to be started at this writing. These results will be published in future *SPSs*.

6. *Generant Tank and Cell Development*

O. F. Keller

In the ALPS system, the generant (hydrazine) which is decomposed to pressurize the propellant tank is stored in the generant tank and cell assembly. A "blowdown" mode of operation is used to expell the generant. This means that the tank is partially filled with generant and then pressurized with gas (nitrogen) which expands as the liquid is withdrawn; no gas is added during the withdrawal process, therefore, the tank pressure decays (from an initial maximum pressure of 1500 psig down to a final minimum of 500 psig in the ALPS system). Entrainment of gas bubbles in the liquid is avoided in the ALPS generant tank design by an expulsion cell which separates the generant from the pressurizing gas.

Reports on this development (*SPS 37-15, 37-17, 37-22, 37-24, 37-30,* and *37-31*, Vol. IV) have presented interim data. At this time the design, fabrication and part of the testing phases have been completed. The following paragraphs relate details of the work on the final version of the generant tank and cell assembly.

a. Design. The generant tank, shown in cross section in Fig. 7 is a sphere, approximately 17 in. in diameter. A safety factor of 2.2 on the maximum working pressure of 1500 psig. The two halves, called the gas-side hemisphere and the liquid-side hemisphere, after the fluid contacted by each, are machined from 6Al 4V titanium forgings which are welded together while kept in alignment by a backup ring. At one pole is an integral boss, housing a small valve through which the pressurizing gas is filled; at the other pole is a small boss to which is welded a small outlet tube through which the generant is filled and expelled. The expulsion cell is a hemisphere made of 0.035-in.-thick ethylene propylene rubber of Shore 80 hardness. It is anchored and sealed to the tank by an O-ring-type lip along its periphery which is clamped near the equator (or girth) of the tank between the backup ring and a sealing surface on the liquid-side hemisphere. When filled with the rated capacity of 1227 in.^{3} of generant, the diaphragm assumes a position which divides the tank into nearly equal volumes and after expulsion, its contour matches that of the lower half (or liquid-side hemisphere) of the tank. A segmented or interrupted rib pattern on the side contacted by the liquid serves to improve expulsion efficiency by preventing the diaphragm from being sealed tightly against the tank wall which might result in part of the liquid being trapped away from the outlet. Small perforated plates cover both of the ports so that the diaphragm cannot be accidentally damaged by extrusion.

b. Extrusion tests. In order to select the proper size holes for the perforated plates, a series of tests was carried out to investigate the ratio of diaphragm thickness to hole diameter as a function of the final generant tank pressure. From the data it appears that for this application, holes with smooth entrances and with diameters less than the diaphragm thickness are not likely to cause damage to the diaphragm. Therefore, the holes in the plates were made 0.040 in. in diameter with the entries rounded to a 0.020-in. radius. These plates prevent extrusion damage even when complete expulsion is achieved at final pressures as high as 700 psig (i.e., 700 psi differential pressure is allowed to develop across the diaphragm).

c. Expulsion efficiency. One of the pertinent performance criteria for any expulsion device is the expulsion

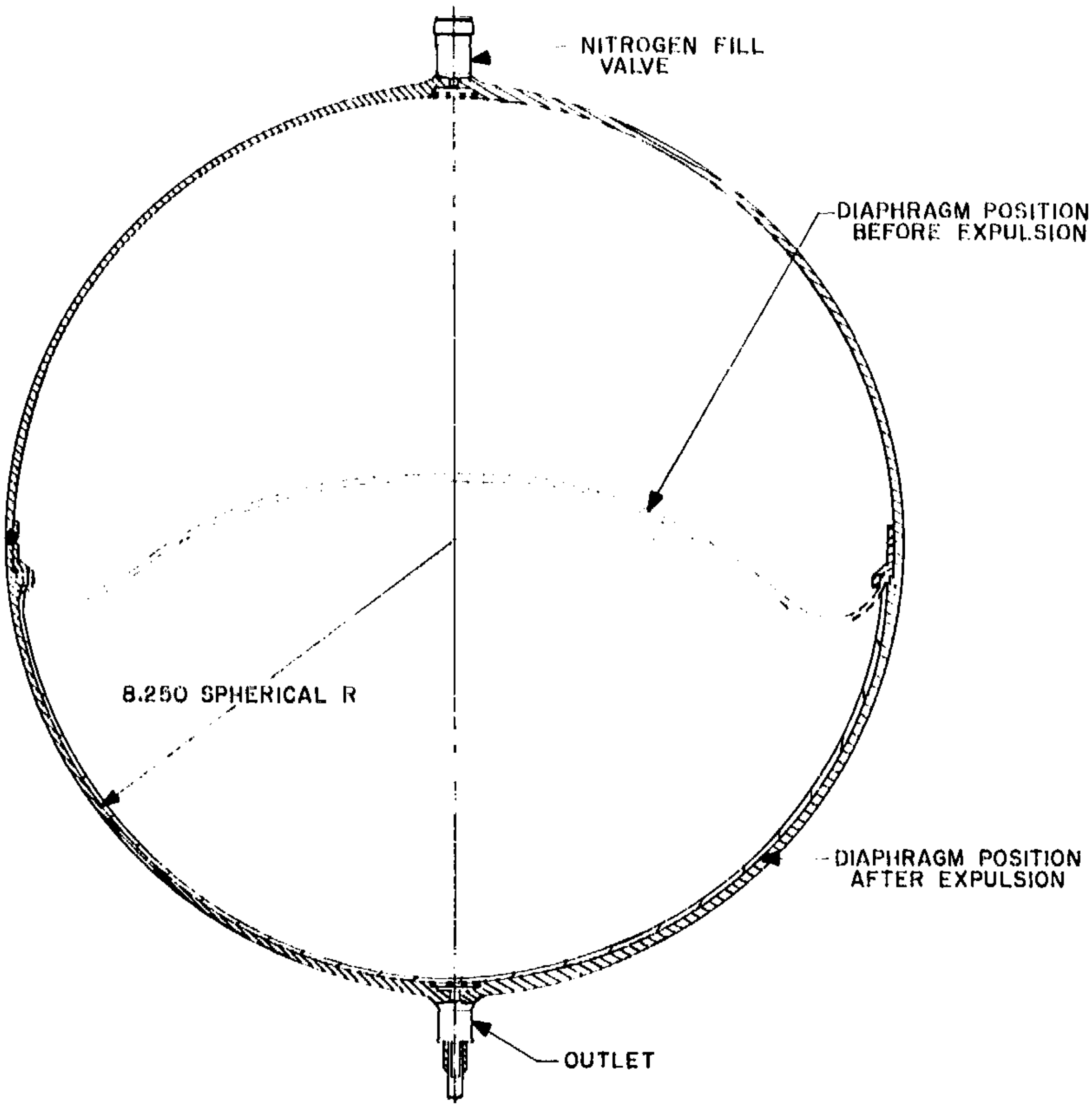


Fig. 7. Cross-section of the ALPS generant tank and cell assembly

efficiency. It is defined as the fraction (often expressed as a percentage) of the liquid put into the device which is expelled when a specified final differential pressure is allowed to develop across the device. Expulsion testing of the tank and cell assemblies has been successfully carried out in the "upright" position at final expulsion pressures of up to 700 psig. Water was the test liquid and nitrogen gas was the pressurant. The trend of the results indicates increasing expulsion efficiency with increasing final tank pressure, as shown in Fig. 8. An expulsion efficiency of 99.85%, or higher, was achieved over a range of final expulsion pressures from 10 psig to 700 psig. (The test data below 100 psig pressure, shown as a dashed line on the graph, were obtained using diaphragms installed in a plexiglass tank.)

d. Diaphragm replacement. One of the obvious disadvantages of the ALPS generant tank and cell design is the difficulty of replacing the diaphragm if it is damaged. After one of the diaphragms was accidentally damaged during proof testing, it was decided that the tank fabricator should attempt to perform the replacement operation. The tank was carefully parted at the girth weld and the backup ring and damaged diaphragm were removed. The edges of the gas-side and liquid-side hemispheres were built up with weld metal and then machined to the original J-groove configuration. To accommodate shrinkage at the girth, the new backup ring was made with a slightly smaller diameter than the original. Then the new diaphragm was bonded to the ring.

The gas-side hemisphere was carefully brought together with the backup ring and diaphragm assembly and tack-welded to it at three places. Next the liquid-side hemisphere was mated with this assembly and water-cooled chill rings were installed. The backup ring was then more substantially tacked to the gas-side hemisphere,

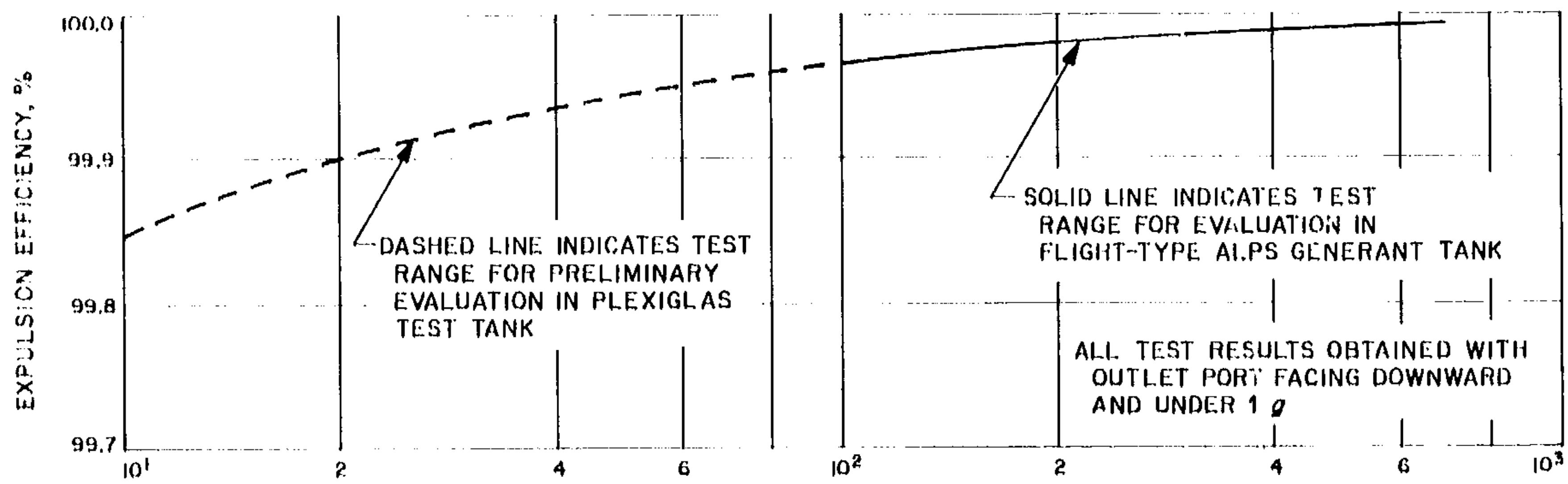


Fig. 8. Expulsion efficiency of the ribbed diaphragm-type ALPS generant cell

using twelve equally spaced tacks approximately ½ in. in length. A preload of 50 lb per lineal inch was applied to the assembly with jack screws, so that the lip of the diaphragm was compressed between the backup ring and the sealing surface on the liquid-side hemisphere. After a preliminary tack welding between the backup ring and liquid-side hemisphere, the welding was completed by several continuous passes which completely filled in the joint.

This repaired tank and cell have successfully passed the proof test, consisting of three cycles of pressurization in which the pressure is maintained at 2250 psig for 5 min during each cycle. A number of water-expulsion tests have also been made with this assembly with no indication of anything except normal performance.

B. Gas-Side Boundary Phenomena

R. W. Rowley

1. Introduction

The development of techniques which can be used to predict local conditions at the wall of an operating liquid rocket engine is essential to the successful application of many new methods of thrust chamber construction. Local heat transfer and local fluid composition are known to influence the erosion rate of ablative materials and of materials used in radiation-cooled thrust chambers. Nonuniform erosion observed in engines fabricated of these materials is recognized to be intimately related to the propellant injection pattern. However, quantitative information on the mechanisms involved in injector related erosion is, in general, not available. As a first step in providing such information, the Gas-Side Boundary Phenomena Project has as an immediate goal the quantitative understanding of the relationship between injector spray properties and local heat-transfer rate to a combustion chamber wall. A discussion of early experimental results and of the rationale which underlies this effort was presented in Refs. 1 and 2. These initial experiments suggested that local heat transfer to a flat plate was related to the local flow rate within the spray. However, heat-transfer rates were lower than desired, and lack of confinement of the burning spray resulted in a combustion field dominated by buoyancy effects. Following completion of these experiments an enclosed combustor was built in an effort to overcome the inherent limitations of the earlier open-air device. This new apparatus is described here and the results of some initial experiments presented.

2. Apparatus

The new apparatus is similar to that used in the earlier experiments except that the spray is now confined within a large, square duct which is open at one end. A single unlike-impinging-doublet injector element is used to produce the spray. The characteristics of this type of element have been investigated extensively at this Laboratory and considerable background information is available. In addition, the unlike doublet is widely used in rocketry because it is a relatively simple element and provides good liquid-phase mixing of fuel and oxidizer.

The propellants used throughout this investigation have been inhibited white-fuming nitric acid (WIFNA) and methanol. When combined at room temperature, these propellants are not hypergolic, although some heat is evolved by liquid-phase reactions. The stoichiometric mixture ratio of 2.30 was used in all tests and ignition was accomplished using a glow coil.

Long tubes (100 L/D) are used as injector orifices to insure reproducible jet properties and, hence, a reproducible spray. Some of the properties of these jets are summarized in Fig. 9. Also shown are the manner in which the burning spray is confined and a simplified interpretation of the spray geometry. The injector is similar to that used in the open-air experiments except for the jet velocities, which have been increased to be more representative of rocket engine practice.

The entire apparatus is best visualized as a simplified, oversize rocket engine from which the nozzle has been omitted. The "combustion chamber" is 20 in. square and approximately 6 ft. long but contains a spray having a total flow rate equivalent to only 300- to 400-lb thrust. The pressure within the duct is essentially atmospheric. Thus, although the physical processes which take place within the burning spray are the same as those which occur within a rocket engine, the rates of some of these processes are reduced by the lower pressure. For example, the rate at which heat is transferred from hot combustion products to the droplets is much lower in the low-density gas generated within the low-pressure device than in a high-pressure rocket engine. The droplets therefore require a longer distance to evaporate and burn, greatly simplifying the problem of making measurements within the spray. Ports located in the sides of the duct allow probes to be inserted through the wall into the spray. Glass windows may be installed in the ports to allow visual observation of the spray or to allow high-speed motion pictures to be taken.

Design of a typical probe used to determine local heat flux is shown in Fig. 10. The probe consists of an instrumented flat plate which is exposed to the burning spray, a protective cover, and a support tube. The heat-transfer rate to the front surface of the plate was calculated using the method of Ref. 3 from temperatures measured at the front and rear surfaces of the plate. A

	OXIDIZER	FUEL
PROPELLANT	WIFNA	CH_3OH
ORIFICE DIAMETER, in	0.182	0.122
FLOW RATE, lb/sec	1.00	0.42
AVERAGE JET VELOCITY, ft/sec	58	103

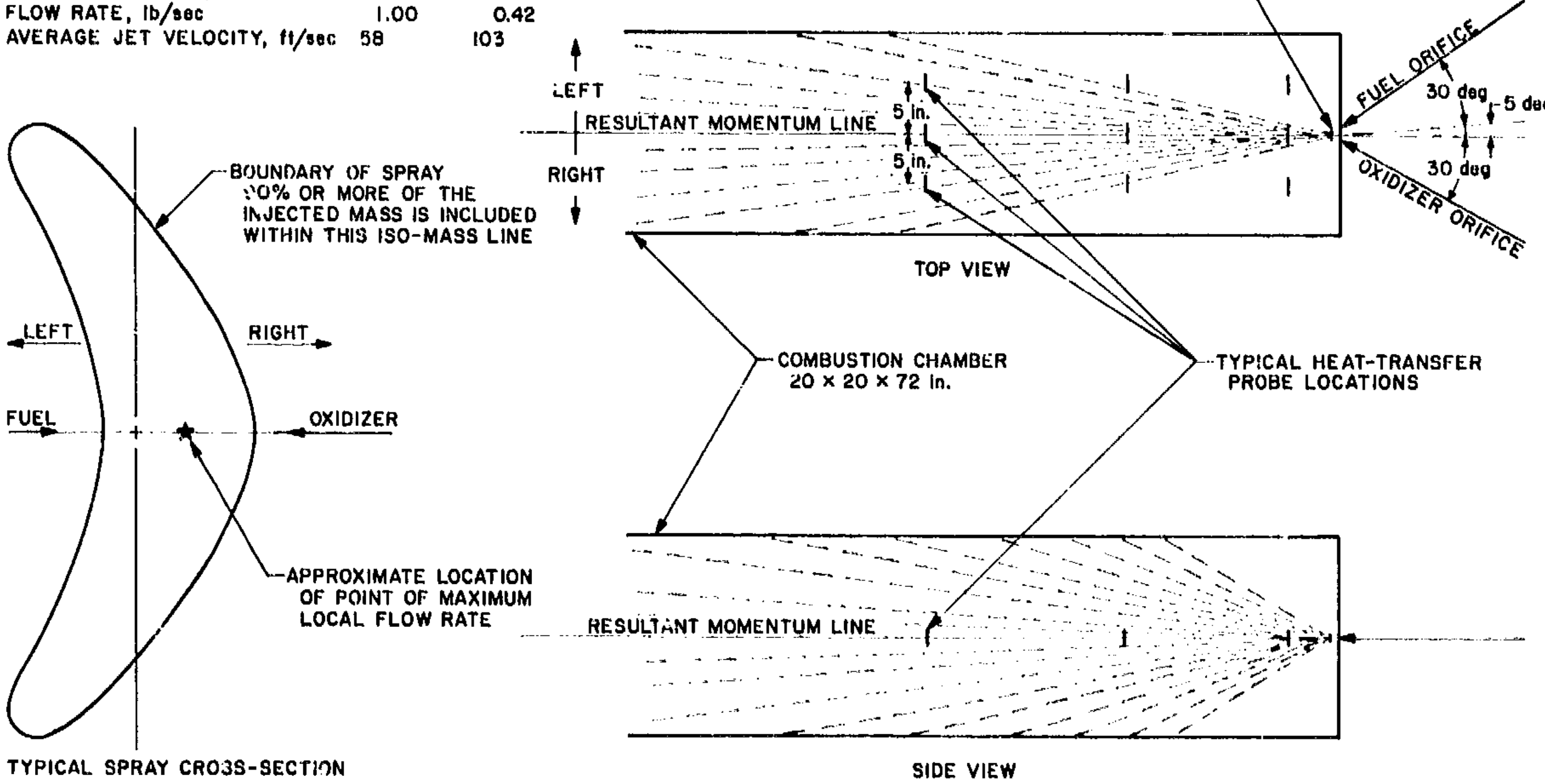


Fig. 9. Configuration of test apparatus and spray

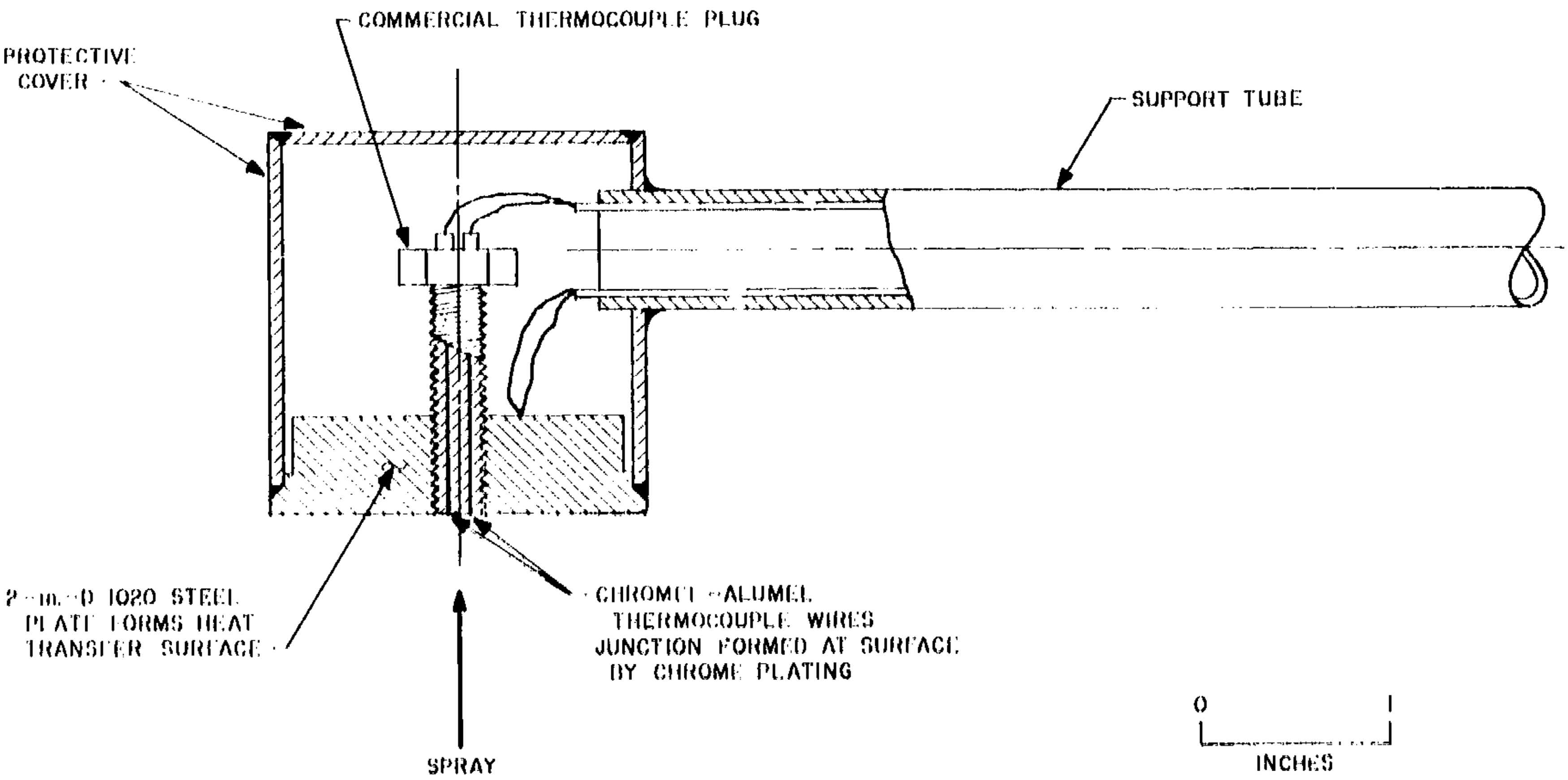


Fig. 10. Heat-transfer probe

commercial Delta-Couple[1] thermocouple plug was used to provide the thermocouple at the center of the front or exposed surface with the thermocouple junction formed by chrome plating. The thermocouple on the rear surface of the plate was attached as near to the center as possible by welding. The protective cover was welded in place to prevent propellants or combustion gases from reaching the thermocouple lead wires. The support tube provided a means of conducting the lead wires through the wall of the duct and also served to locate the probe within the spray.

For the tests reported here, the probe was oriented with the face perpendicular to the resultant momentum vector of the spray. Thus, when the probe was located along the resultant momentum line (i.e., the longitudinal centerline of the duct) the center of the probe represented a stagnation point. When the probe was displaced from this centerline, impingement of the spray was no longer perpendicular to the face and a small relative velocity existed at the thermocouple location. However, limited testing in the open-air experiment indicated that small deviations from true stagnation flow did not result in measurable differences in heat flux. Based on this information it is believed that the deviations from stagnation flow in the experiments reported here do not significantly affect the magnitude of heat flux observed.

[1]Advanced Technology Laboratories, Mountain View, California

3. Results

Fig. 11(a) presents results from this new experiment, while Fig. 11(b) shows the results from the open-air device, as reported in Ref. 2. All data are presented in terms of heat-transfer rate as a function of distance downstream from the impingement point at positions either on the centerline or 5 in. to either side, as shown in Fig. 9. Heat transfer rate was calculated at a time 2 sec after initiation of propellant flows. The increase in magnitude of heat flux at the same location in the spray is striking, amounting typically to a factor of 5 higher in the enclosed experiment than in the open service.

As was discussed in Ref. 2, the magnitude of heat flux in the open-air spray bore a relationship to the local mass flow rate. This was also the case in the present experiment. By examining Fig. 9, and noting that the point of maximum local mass flow rate lies to the right of the resultant momentum line (i.e., the centerline of the duct),

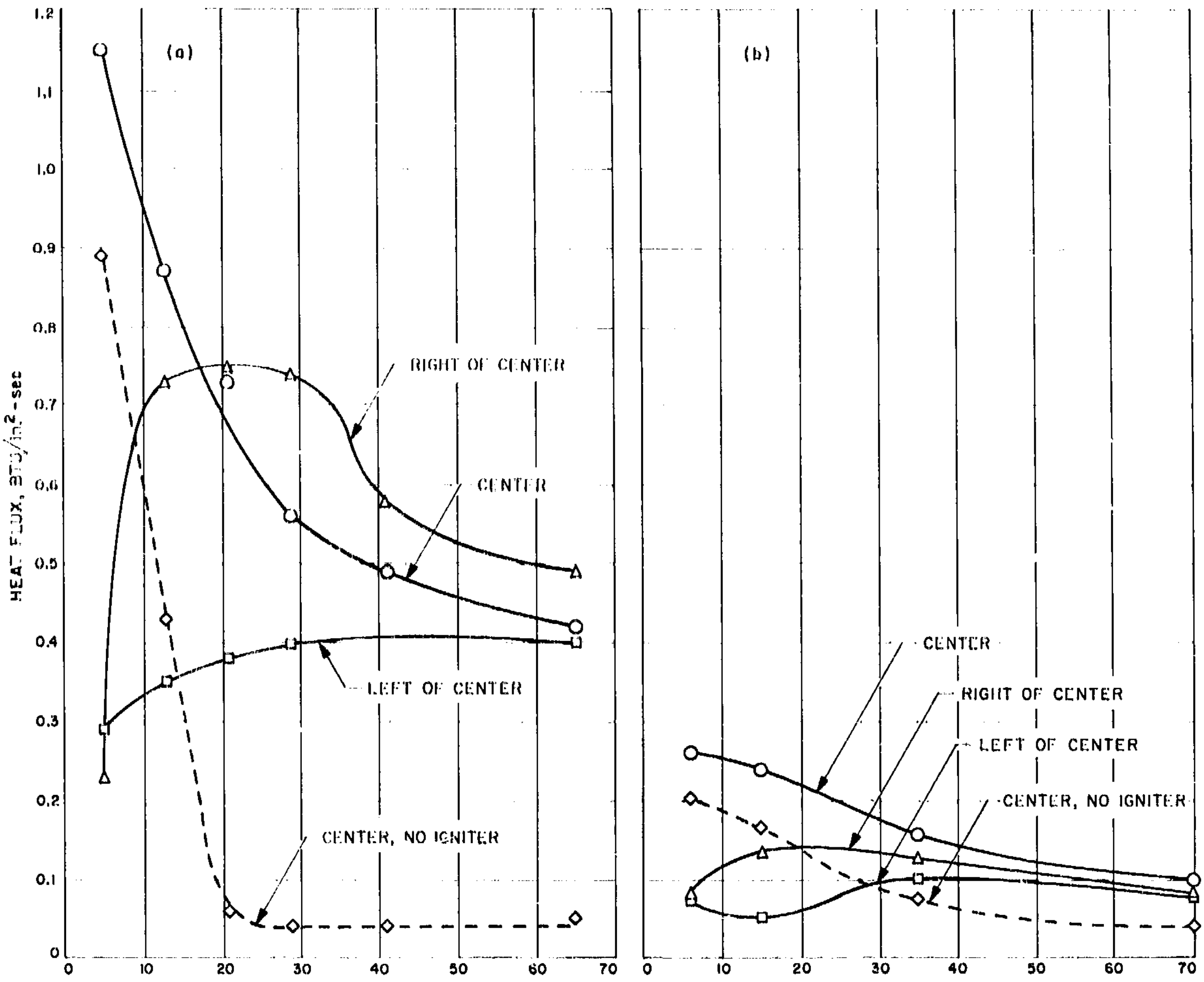


Fig. 11. Heat-transfer rate to flat probe (a) enclosed experiment (b) open-air experiment

the local heat-transfer rate is seen to vary as the local mass flow rate. Bear in mind that: (1) as distance from impingement increases, the cross-section of the spray retains its shape but increases in size; (2) local flow rate decreases as the square of the distance from impingement and; (3) the proportion of hot combustion products in the spray increases with distance from impingement. As an example, consider the heat-transfer rate to the flat plate when the plate was oriented along the duct centerline. Heat flux was highest when the plate was close to the impingement point and decreased with increasing distance. The curve labeled "no igniter" indicates the heat-transfer rate from a spray which was not burning, the heat flux resulting entirely from droplets which had been heated by the liquid phase reactions between fuel and oxidizer. The heat flux measured in this nonburning case was inversely proportional to the square of the distance from impingement, indicating a first-power dependence on local mass flow rate. The difference in heat flux between the burning spray and the nonburning spray is a measure of the contribution of the hot, gaseous combustion products to the total amount of heat transferred to the plate.

The dependence of local heat-transfer rate on local mass flow rate was evident throughout the spray. Along the line 5 in. to the right of the centerline, the heat flux: (1) was quite low when the probe was close to the impingement point; (2) increased rapidly to a maximum at 20-25 in. from impingement; and (3) decreased with further increase in distance. As may be seen from Fig. 9, these locations correspond to exposure of the probe to a mass flow rate which was respectively: (1) quite low, since the probe was essentially outside the boundary of the spray when located close to the impingement point; (2) increasing as the probe was exposed to the part of the spray cross-section having the highest local mass flow rate; and (3) decreasing as the distance from impingement increased and local flow rate decreased.

At the position 5 in. to the left of the centerline, the probe was outside the boundary of the spray when close to the impingement point, and, so indicated a low heat flux. With increasing distance, the probe passed into the "edge" of the spray, and heat flux increased. However, the left side of the spray represents a region of relatively low local flow rate, and heat flux was always lower than in the center or right portions.

4. Discussion

It may be concluded that by containing the burning spray and excluding the surrounding air, the desired increase in droplet consumption rate and local heat-transfer rate was achieved. The effect of air as a diluent or as a competitive oxidizer was more significant than had been estimated previously, and the increase in heat flux by a factor of 5, even in the center of the spray, was unexpected.

Although the limited data presented here indicate that local flow rate is a significant determining factor in local heat flux, no quantitative relationship has yet been determined. However, work is in progress, using nonreactive fluid techniques, to measure the local flow rates within the spray used in these experiments. Additional heat-transfer measurements are also being made. It is intended that a quantitative correlation be derived to relate local heat-transfer rate to local spray properties, including, as a minimum, local flow rate and a measure of the degree of completion of reaction.

References

1. Rowley, R. W., "Gas Side Boundary Phenomena," *SPS 37-26*, Vol. IV, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, April 30, 1964, pp. 118–123.

2. Rowley, R. W., "Gas Side Boundary Phenomena," *SPS 37-28*, Vol. IV, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, August 31, 1964, pp. 71–76.

3. Powell, W. B., Howell, G. W., and Irving, J. P., *A Method for the Determination of Local Transient Heat Flux in Uncooled Rocket Motors*, Technical Report No. 32-257, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, July 1, 1962.

XVI. Lunar and Planetary Instruments

A. Slow-Scan Vidicon Sterilization Tests

L. R. Baker

1. Introduction

Because of biological constraints associated with landed planetary spacecraft, requirements have been generated for sterilization compatibility tests on science instruments scheduled for advanced planetary exploration. In imaging systems scheduled for such missions, the image sensor will be required to pass the sterilization compatibility test, which includes three 36-hr 145°C-temperature soaks. At the present time, the image sensor used in spacecraft imaging systems is the vidicon.

The vidicon is a rugged, reliable, fairly sensitive image sensor that has been used extensively in commercial and industrial television systems for over 14 yr. Because of this background of operating requirements in near-human environments, no attempt was made to design the vidicon to survive extremely high temperatures until relatively recently. However, in the space exploration program, storage and, in some cases, operation of vidicons at high temperatures are required.

2. Vidicon High-Temperature Problems

The problems of designing a vidicon to withstand high temperature are generally associated with the photosensitive layer called the photoconductor, and with the vacuum seal between the vidicon faceplate, on which the photoconductor is vacuum-deposited, and the envelope. This seal is a cold solder seal usually made with the metal indium. It was the development of this sealing technique that brought the vidicon into the role of prominence it now enjoys. However, indium in its pure state melts at a temperature of 155°C, which is very close to the sterilization temperature of 145 ±2°C.

The storage of photoconductors at high temperatures can cause severe degradation of image quality. The photoconductor is a thin film of semiconductor material which has been deposited on the glass substrate in a hard vacuum. Under high temperature conditions, this material can crystallize, vaporize, or it can change characteristics such as to severely degrade the image quality. This degradation is characterized by the appearance of spots or loss of resolution, and can be accompanied by a change in the operating characteristics of the tube.

3. Vidicon High-Temperature Testing

In an attempt to better understand the problems associated with vidicons at high temperatures, three vidicons that were procured from two different vendors under the *Ranger* program were subjected to high temperature testing. Vidicons 1 and 2 incorporated a proprietary photoconductor, but it is known to be a derivative of selenium. Vidicon 3, on the other hand, incorporated a photoconductor of antimony sulfide oxysulfide, commonly referred to as ASOS.

The test method was to operate and optimize the vidicon in a camera tube test set and measure the pertinent operating voltages and currents, then photograph the image from the vidicon as displayed on a television monitor. The tube was then subjected to 6 hr of temperature soaking, then retested in the camera tube test set using the same operating voltages and currents.

The voltages and currents measured were:

(1) G_1 voltage.

(2) G_2 voltage; usually 300 to 350 v dc

(3) G_3, G_4 voltages (mesh); usually 950 to 1000 v dc.

(4) Filament voltage.

(5) Filament current.

(6) Target (signal electrode) voltage.

(7) Signal current from target.

(8) Focus coil current.

The illumination for the vidicon was derived from a collimator associated with the camera tube test set and was held at 0.1 ft-c at 2870°K. Table 1 is a brief summary of the test results of the three vidicons.

Table 1. Vidicon test results

Test	Vidicon 1	Vidicon 2	Vidicon 3
Pretest	Few white spots, fair resolution	Few white spots, good resolution	No spots, fair resolution
Post 100°C	Many white spots, poor resolution	Many white spots, good resolution	No spots, fair resolution (no change in quality)
Post 110°C	—[a]	Large single white spot, poor resolution	—[a]
Post 115°C	—[a]	—[a]	Tube became gassy; would not operate
Post 120°C	—[a]	No video, dark current increased, photoconductor partially migrated	—[a]
Post 125°C	Photoconductor migrated, mesh "exploded"	Photoconductor completely migrated, mesh "warped"	—[a]
Post 155°C	Faceplate fell away from envelope	—[a]	—[a]

[a]No test on this tube at this temperature.

The photographs of the test on Vidicon 1 are illustrated in Figs. 1–5. After failure of Vidicon 1 at 125°C, it was decided to test the faceplate-to-envelope seal (indium) on this vidicon to destruction. (The indium used in the cold solder seal has a melting point of 155°C.) The vidicon was subjected to increasing temperature increments of 2°C until the seal failed. After each temperature increment, the transconductance of the tube was

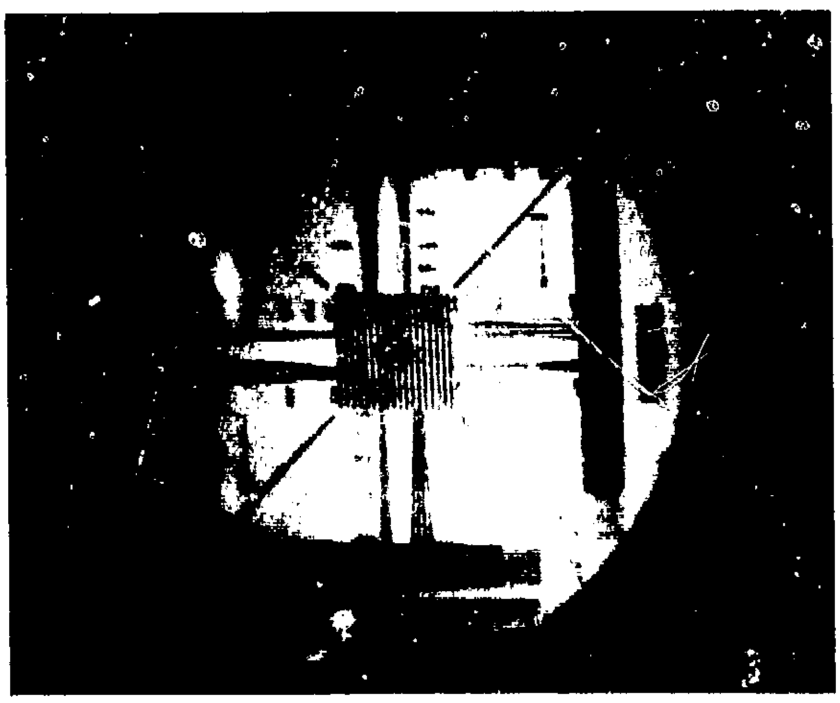
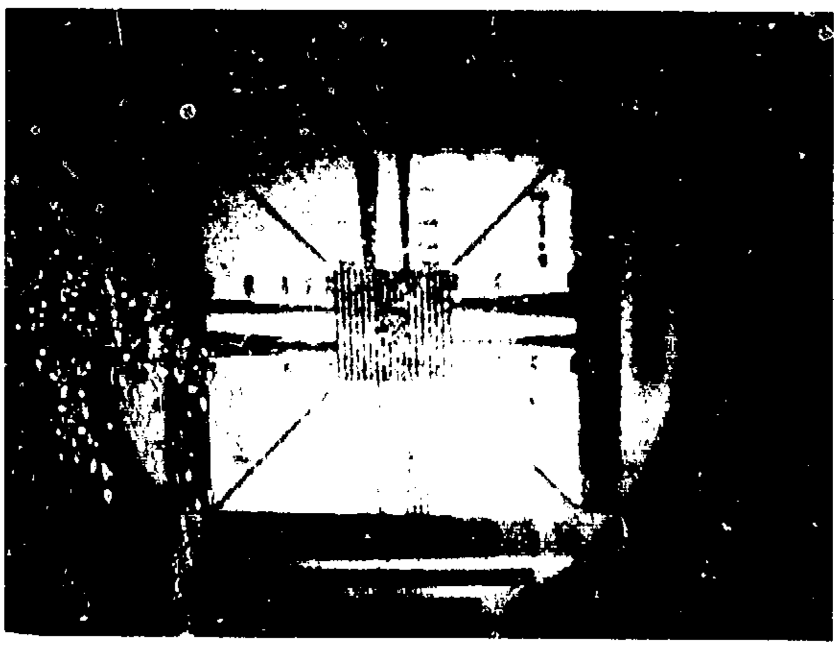


Fig. 1. Vidicon 1 pretest RETMA

F'g. 2. Vidicon 1 pretest black field

Fig. 4. Vidicon I post 100°C black field



Fi ,. 3. Vidicon 1 post 100°C RETMA

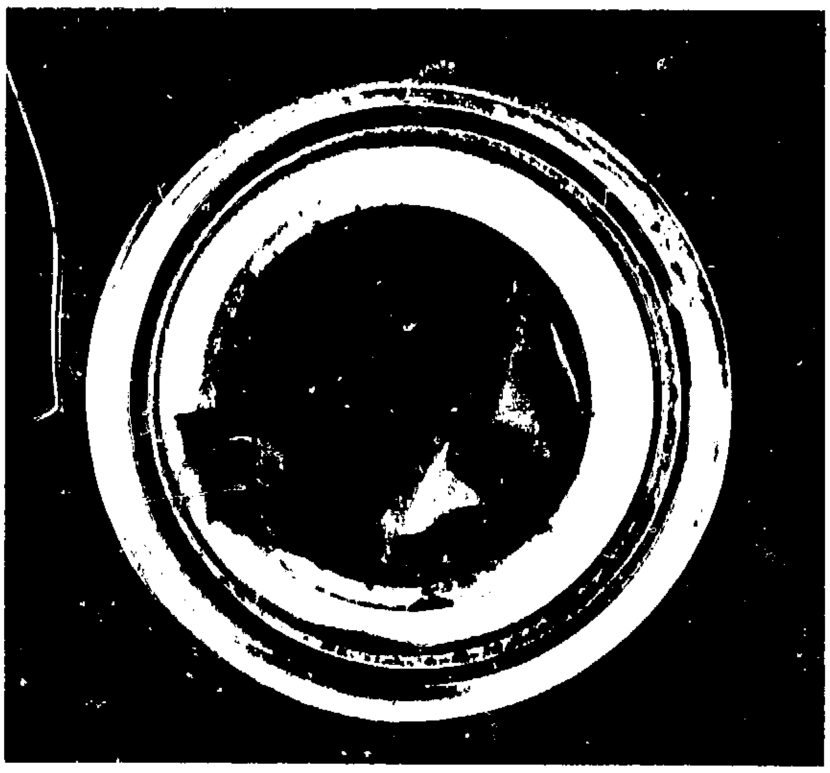


Fig. 5. Vidicon 1 faceplate and mesh

measured using Grid 2 as the anode. The transconductance did not change significantly from pretest to the post 153°C test, as shown in Fig. 6. At 155°C, however, the indium melted, causing the faceplate to fall away from the envelope, as shown in Fig. 7.

The photographs of the test of Vidicon 2 are illustrated in Figs. 8–15. During the 120°C bake, the target material started to migrate; during the 125°C period, the target material migrated completely and the mesh "warped," which is very evident in the television monitor photographs (Figs. 14 and 15). No further testing was done on this tube.

Photographs of the test on Vidicon 3 are shown in Figs. 16–19. This vidicon has different photoconductive material and withstood the 100°C sustained temperature better than either Tubes 1 or 2. However, the tube became gassy and stopped operating during the 115°C temperature test; the test was, therefore, concluded. An examination of the photoconductor and mesh did not reveal either photoconductor migration or mesh deformation. Since the temperature was not increased beyond the 115°C point, a comparison between Vidicon 3 and Vidicons 1 and 2 in terms of the physical parameters is not totally possible at this time.

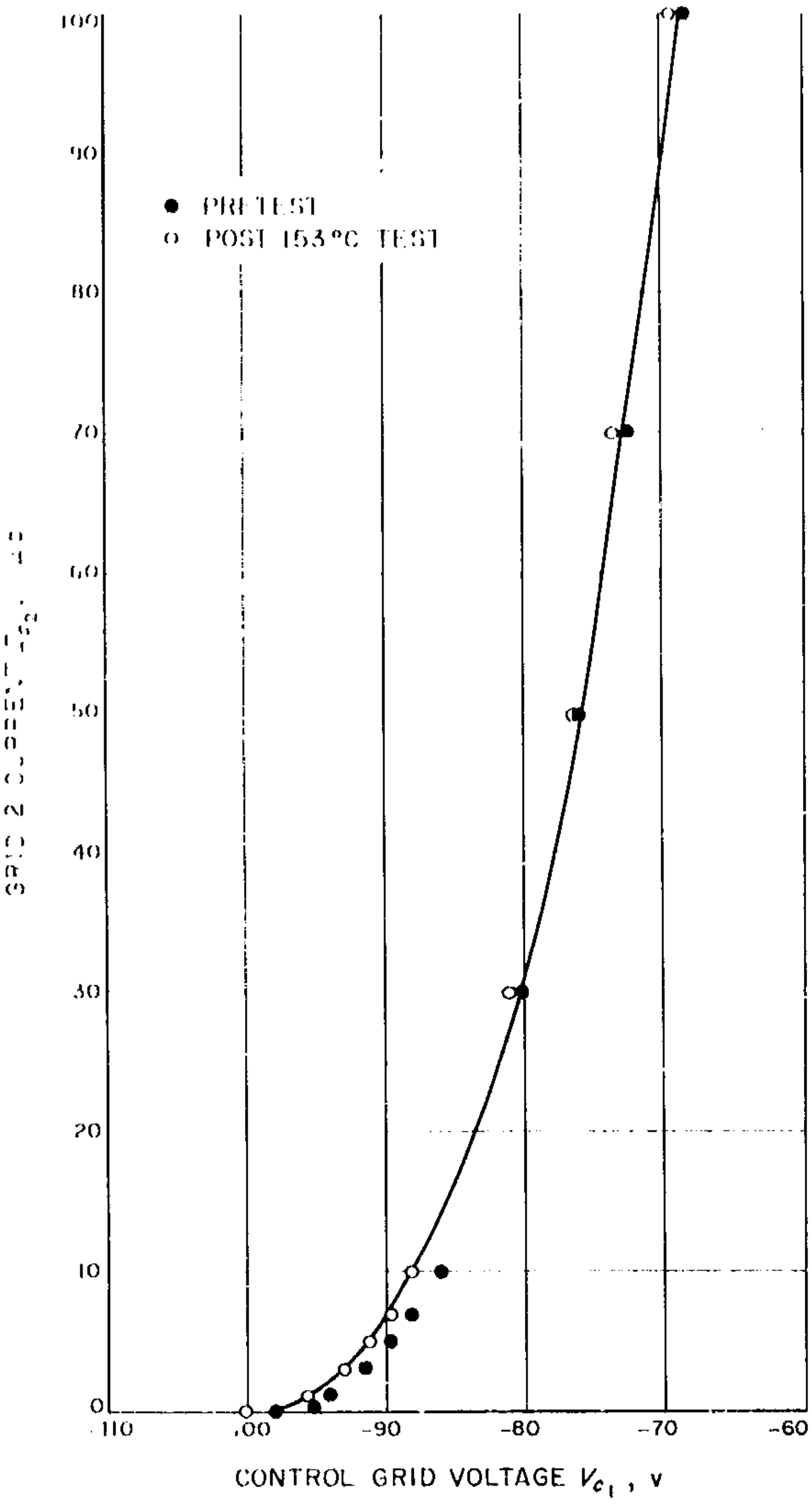


Fig. 6. Vidicon 1 transconductance comparison

Fig. 7. Vidicon 1 faceplate seal melted

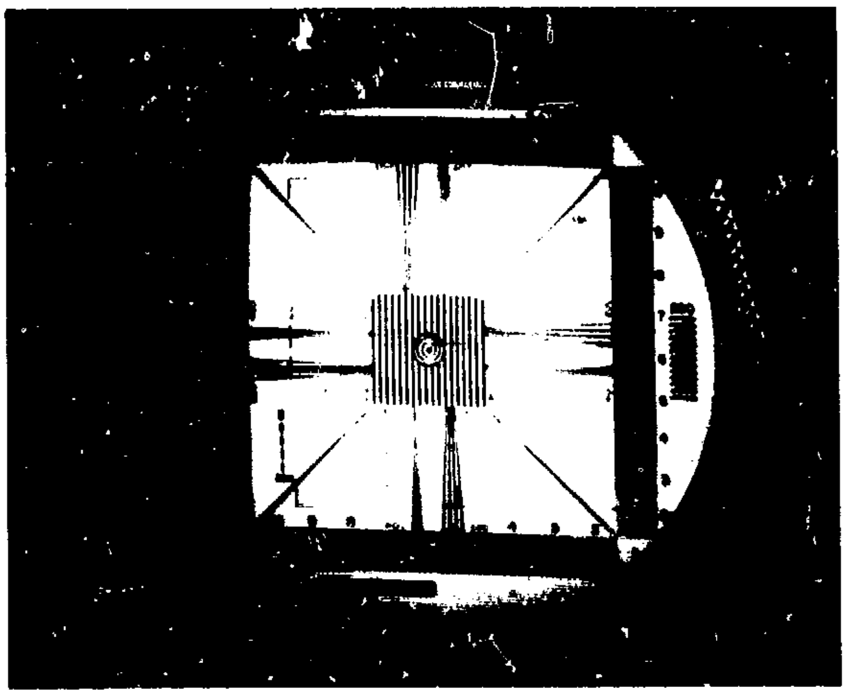


Fig. 8. Vidicon 2 pretest RETMA

4. Conclusions

The tests conducted so far indicate that much more work is required on vidicons to survive the sterilization compatibility test. The photoconductors on Vidicons 1 and 2 evaporated and migrated to other areas of the tube between 120 and 125°C. The mesh failures on these tubes may have been caused or helped by the photoconductor evaporation, but no exact cause is known at this writing.

Vidicon 3 survived the 100°C sustained temperature with no degradation. It is unfortunate that the tube became gassy since no other vidicon of this type was available at the time of the test activity.

A program is presently under way to design a vidicon which will be capable of surviving the sterilization compatibility test. With a sterilizable image sensor, breadboard instruments can be designed and built with confidence, knowing that future programs requiring sterilization may be able to carry image sensors of the vidicon type.

Fig. 9. Vidicon 2 pretest black field

Fig. 11. Vidicon 2 post 100°C black field

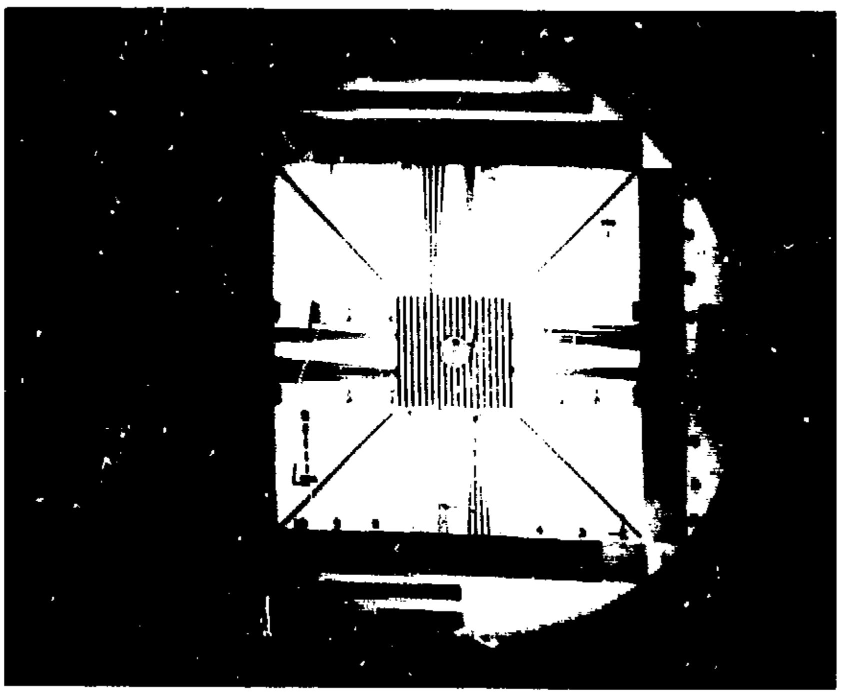


Fig. 10. Vidicon 2 post 100°C RETMA

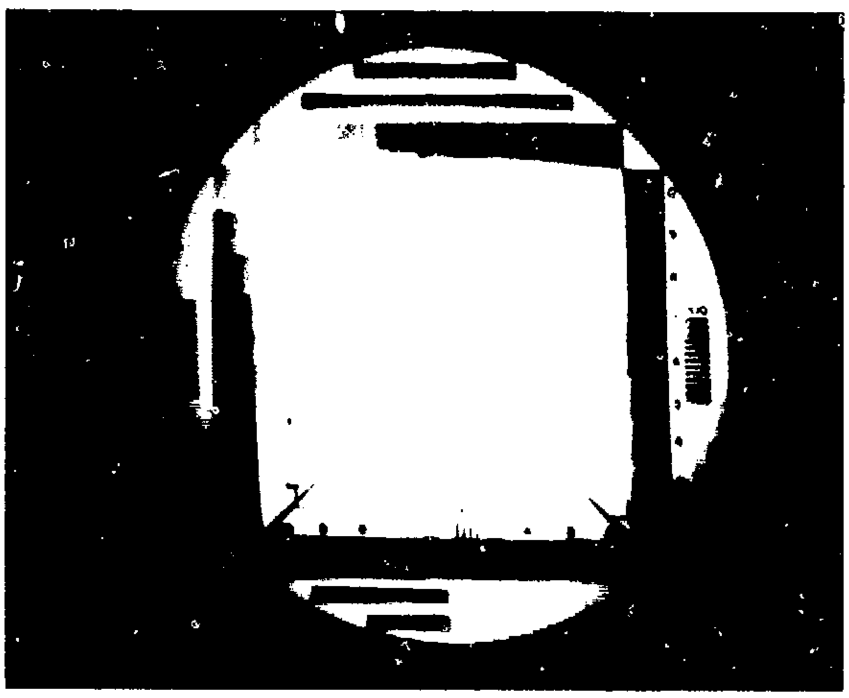


Fig. 12. Vidicon 2 post 110°C RETMA

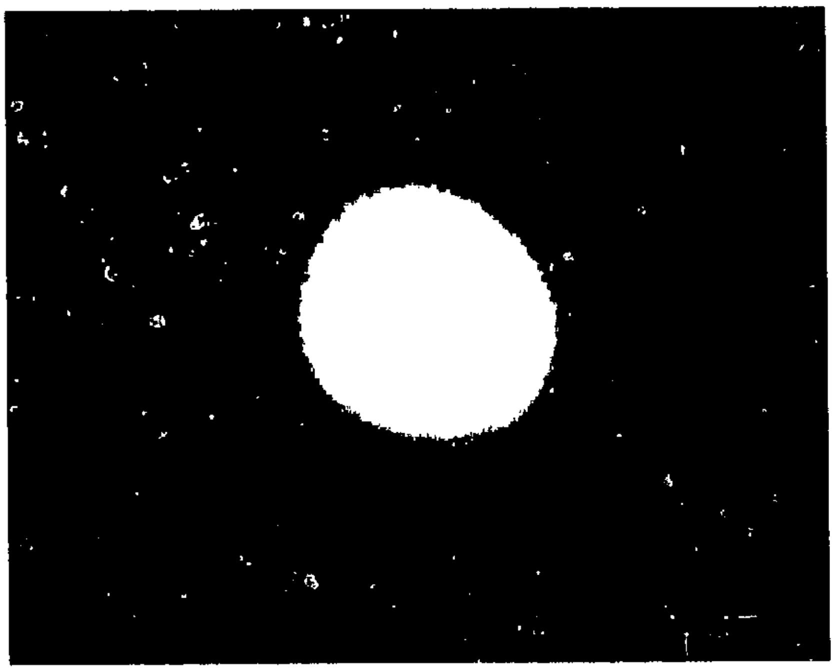


Fig. 13. Vidicon 2 post 110°C black field

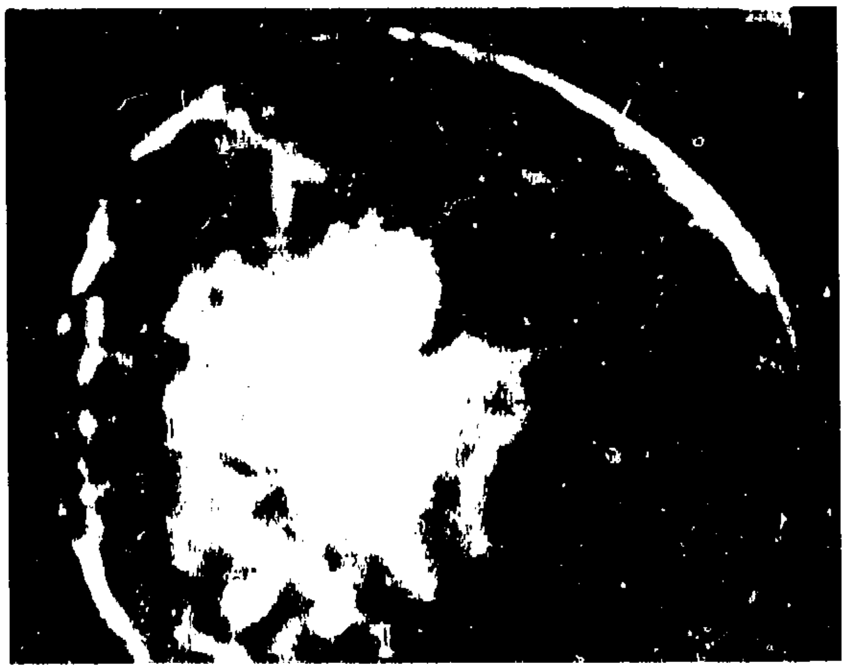


Fig. 15. Vidicon 2 post 125°C

Fig. 14. \ dicon 2 post 120°C

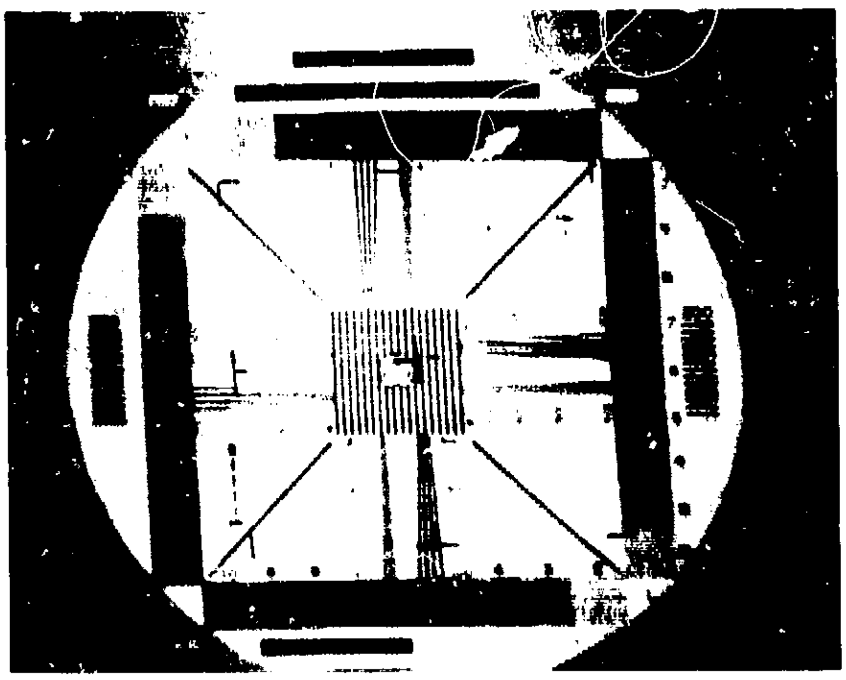


Fig. 16. Vidicon 3 pretest RETMA

Fig. 17. Vidicon 3 pretest black field

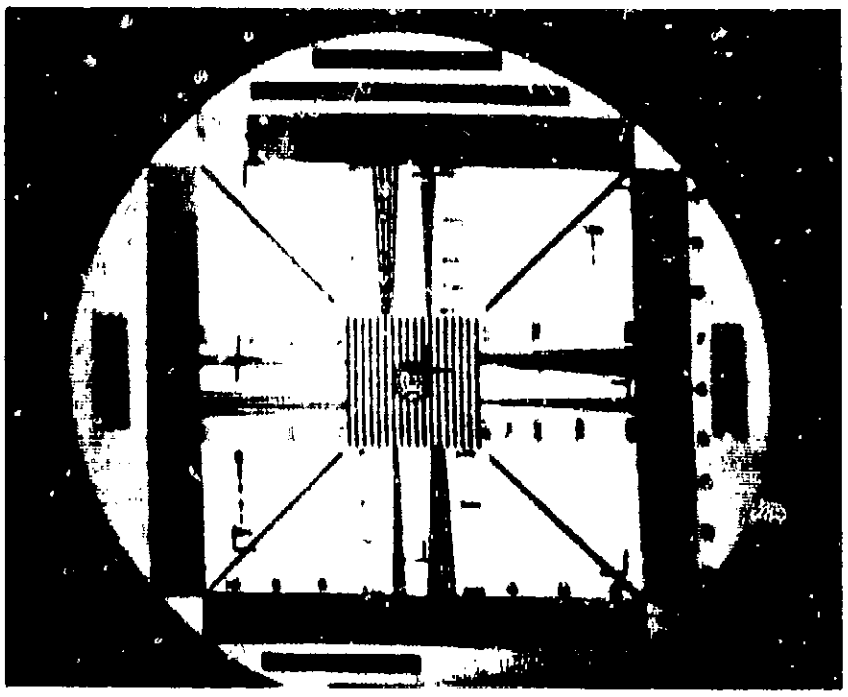


Fig. 18. Vidicon 3 post 100°C RETMA

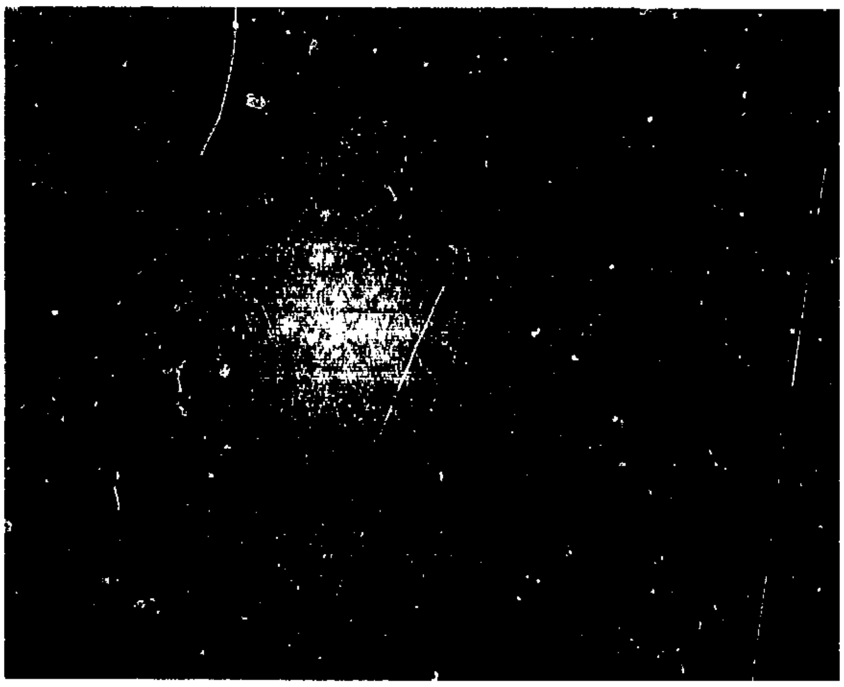


Fig. 19. Vidicon 3 post 100°C black field

XVII. Space Instruments

A. A Camera-Film Processor for Planetary Reconnaissance and Data Storage

A. W. Spitzak

There are two types of systems immediately available for use in space reconnaissance, the vidicon TV system and systems using silver halide films. This statement is based on the fact that these are the only types of systems used extensively in space up to the present time.

Silver halide systems have advantages in large-area reconnaissance at planetary distances. Here, the inherent storage of the film allows us to bypass the problem of mating a TV camera to a tape recorder to obtain the necessary reduction in bit rate commensurable with the communication system. This, of course, is a problem faced by the *Mariner C* system (Mars 1964 mission). Low communication rates will be with us for some time, but more area coverage is going to be demanded in later missions. In general, the rate at which the picture frames will be taken cannot be made slow enough to be used directly in the communications system. This is no problem in a silver halide system since there is no interplay between the read-in, read-out speeds. The read-in speed can be fixed solely by the desire to take a series of pictures at a given interval of time, and the read-out speed can be fixed to be compatible with the communication system.

The full capability of silver halide film can only be exploited in a suitably designed system. Added advantages can be realized by utilizing one or more of the following features:

(1) Non-reconnaissance data storage as a replacement or supplement for a spacecraft tape recorder.

(2) Large-area low-resolution scanning with high resolution capability still available.

(3) Availability of film in many format sizes.

Although the film is being considered primarily because of reconnaissance data storage capability, there are other instruments postulated for Martian spacecraft use that produce larger amounts of information than the communication system can handle in real time. The large storage capability of film could be used to simplify this problem. Again, the read-out speed could be made independent of the recording speed and many simultaneous recordings could be accomplished at the same time.

In communicating the reconnaissance information, a particular frame could be scanned with a larger spot size and faster rate, to get a gross look at the planet. In conjunction with this large spot scan, a smaller spot size scan of the same frame might be made in the spacecraft, and a number representing, say, the average number per line of edge gradients exceeding a certain amount, sent along with the large spot size scanned picture. The number and picture might be used to decide which of the many pictures available is first chosen for complete small spot size scanning.

If system weight were not too restrictive, a film scan system capable of 80–100 TV lines/mm is available. However, a practical system of reasonable weight in an earlier planetary mission would be closer to 40 TV lines/mm. Potential for higher resolution in a film system is there and is limited by allowable weight. TV vidicons have an upper practical limit of 40 TV lines/mm, limited by available scanning beam diameter.

All of these above-stated features make the investigation of the problems associated with silver halide film systems very worthwhile. These problems can be considered in three principal areas:

(1) Film exposure, transport, and processing mechanization.

(2) Film read-out mechanisms.

(3) Radiation damage versus sensitivity of films and damage to chemical developers from radiation.

In order to investigate the problems of film exposure, transport, and processing, a system was constructed by Fairchild Space and Defense Systems for laboratory use. Basically, the system is a framing type camera and a film processor in a self-contained package (Figs. 1–3). One hundred feet of 35-mm film is carried. The primary task of the system is to demonstrate the capability of exposing and processing after 8 months of unattended operation. With present film technology, the film, sensitized on Earth, along with the necessary chemicals for processing, must be transported to the vicinity of Mars.

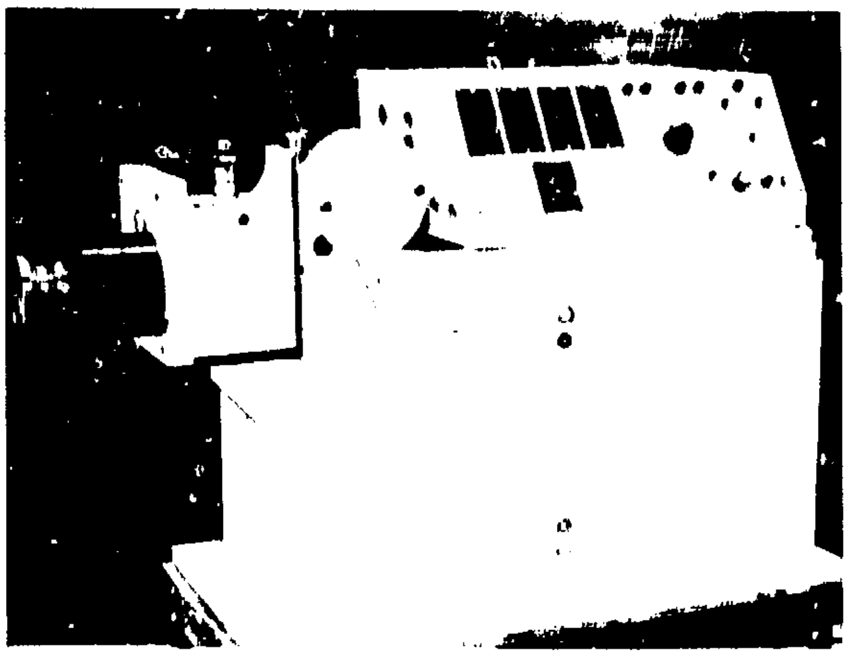


Fig. 1. Camera-film processor (oblique view)

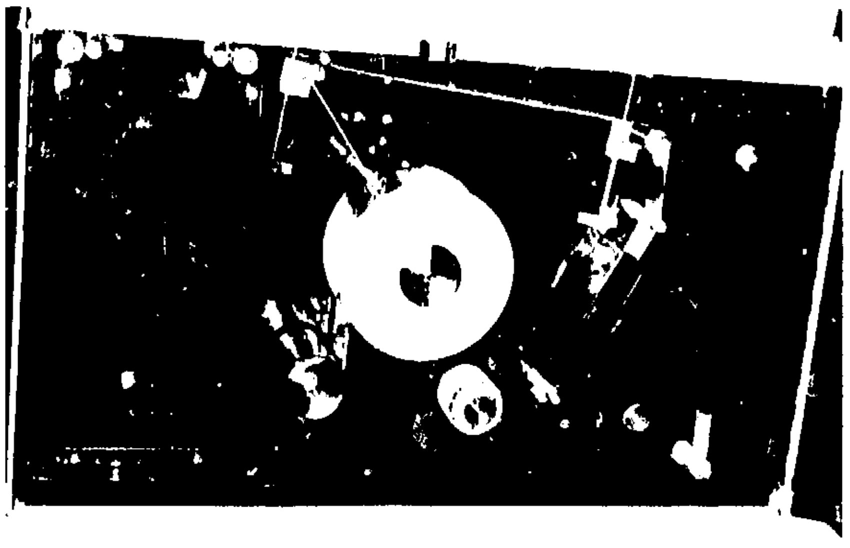


Fig. 2. Camera-film processor (after camera operation with processor; mode on)

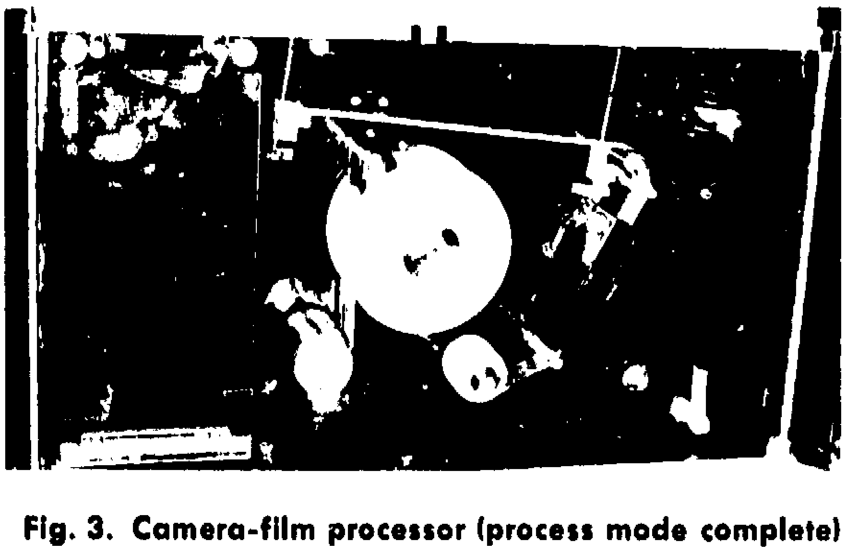


Fig. 3. Camera-film processor (process mode complete)

Film processing is still best done by wet chemicals. The system built by Fairchild uses a presoaked web process to avoid the complexities of a pump system and fluid storage. The chemicals in the web are a monobath mixture containing both the necessary developing and fixing solutions. Timing problems are a minimum with monobaths because they are formulated to carry the processing to completion as long as a minimum time is exceeded. Here, that time is about 3 min.

The chemical-carrying web in this system is a microporous-structured inert material of unplasticized polyvinyl chloride about 0.008 in. thick. About 80% of this thickness is void volume in which the processing chemicals are held until brought in contact with the film emulsion, which has a high affinity for the water-based chemicals. The web material has good wet strength and does not compress easily (will not drip). It is a patented Fairchild product with the trade name Poromat. Impregnated webs have been frozen at 0°F, thawed, and used satisfactorily.

The Kodak films for which monobaths have been specifically formulated are SO-243 (exposure index of 1.6) and 5240 (exposure index of 40) (Figs. 4 and 5). The low film speeds were chosen principally to minimize the radiation damage to film; however, in the overall system considerations, the grain of the film can be a principal contribution to the noise in the system and these low-speed, fine-grain films may have system advantages that counterbalance their lower sensitivities.

The camera and its associated controls have the capability to expose a number of frames at various rates. The frame size is 25 by 70 mm, and up to 12 ft of film can be exposed and processed at any one time.

Film exposed in the camera is dumped into a free slack box until all exposures are complete. The process mode is then started and the saturated web is brought in contact with the film, transported with the film for the necessary processing time, and then separated. The film is now slightly damp to the touch and is carried through a desiccant drying station before being wound on a cassette. The film is now in a state completely prepared for electronic scanning.

Since there is no washing of the film, as is experienced in a darkroom, and the processed film is to be used in a film scanner which will measure the transmitted light through the film, no processing chemical blotches should be experienced on the dried film. The monobaths formulated for the films to be used are very clean and the films do not show this effect.

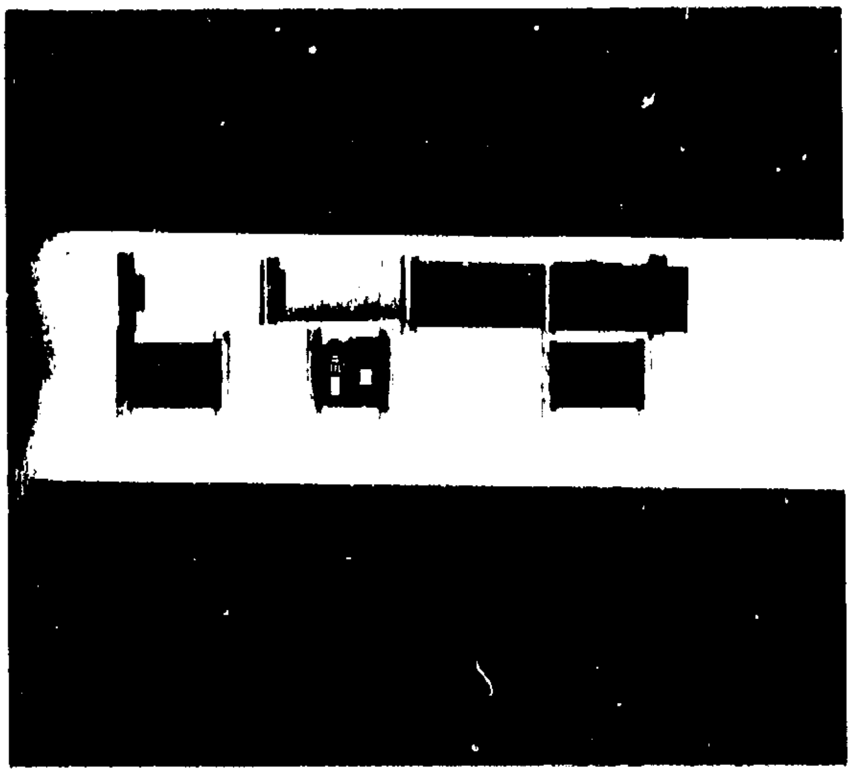


Fig. 4. Kodak film SO-243 (35mm) as processed in camera-film processor showing perforations for size comparison

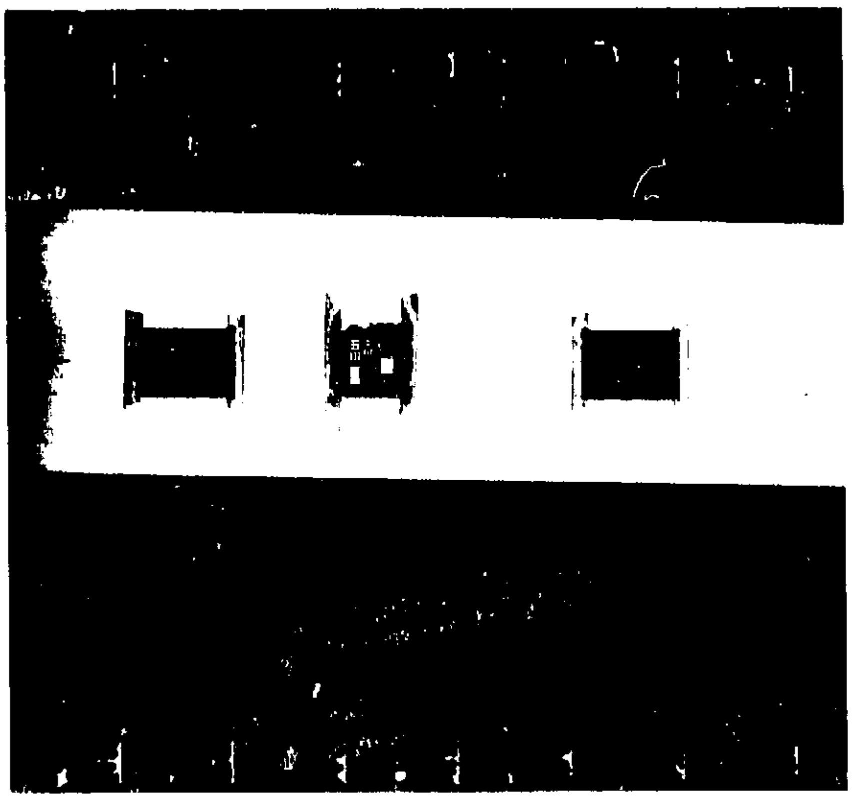


Fig. 5. Kodak film 5240 (35mm) as processed in camera-film processor showing perforations for size comparison

The impregnated web can be made useless by two principal causes: evaporation of the chemicals and oxidation of the chemicals. The impregnated web is protected from these two environments by sealing it in a sandwiching of two pieces of aluminized mylar. The particular environments at which this operation is optimally carried out to obtain 8 months of useful life is under study. Additional studies are also being conducted to determine the best packaging materials. At the time of use, the sandwich will be cut open and the mylar peeled back and rolled up. Only as much web as is needed for development of a film segment will be brought in contact with the film; the rest of the impregnated web remains protected from any hostile environment.

Poromat was considered to have the best capability of meeting the 8-month storage requirement. Another web material considered was the Kodak product BiMet. This material is not inert and is essentially a gelatin-base carrier which, at the present time, cannot exist in its impregnated state for longer than 4 months. Its use in this system would require internal spacecraft impregnation and it would be just as well to impregnate the film itself with the processing chemicals.

Extensive environmental studies will be undertaken with the camera processor to determine the restrictions on the processing environment, and the long-term required storage environment, especially as it pertains to the film. From this data, the required spacecraft environment will be bounded.

B. Film Scan System Using a Solid-State Light Source and Light Detector

A. W. Spitzak

The system described in the previous article[1] will require a scanner system to convert the film density to electrical signals for data transmission. Some of the design restrictions on the scanner to be used are discussed below.

[1] "A Camera-Film Processor for Planetary Reconnaissance and Data Storage," pp. 174–177.

A film scanner should operate directly into the communication system at the requisite communication bandwidth. If faster scanning rates were used than could be accepted by the communication system, buffer data storage would be required. In the device to be described, the bandwidth considered is less than 20 cps.

The scanning spot size is determined by considerations of the film being used. The scanning spot size should be compatible with the film grain noise and camera-film system modulation transfer restrictions, and the signal-to-noise ratio of the scan system should allow good utilization of the dynamic range of the scene as imaged on the film.

With these restrictions, a scanner development was undertaken. The device maintains a rigid alignment between the light source optics and light sensor (Fig. 6). The film is transported relative to it to obtain line and frame scan. The innovations in the device and the reason for this article are the utilization of a P-N junction diode as a light emitter and a silicon transistor as light

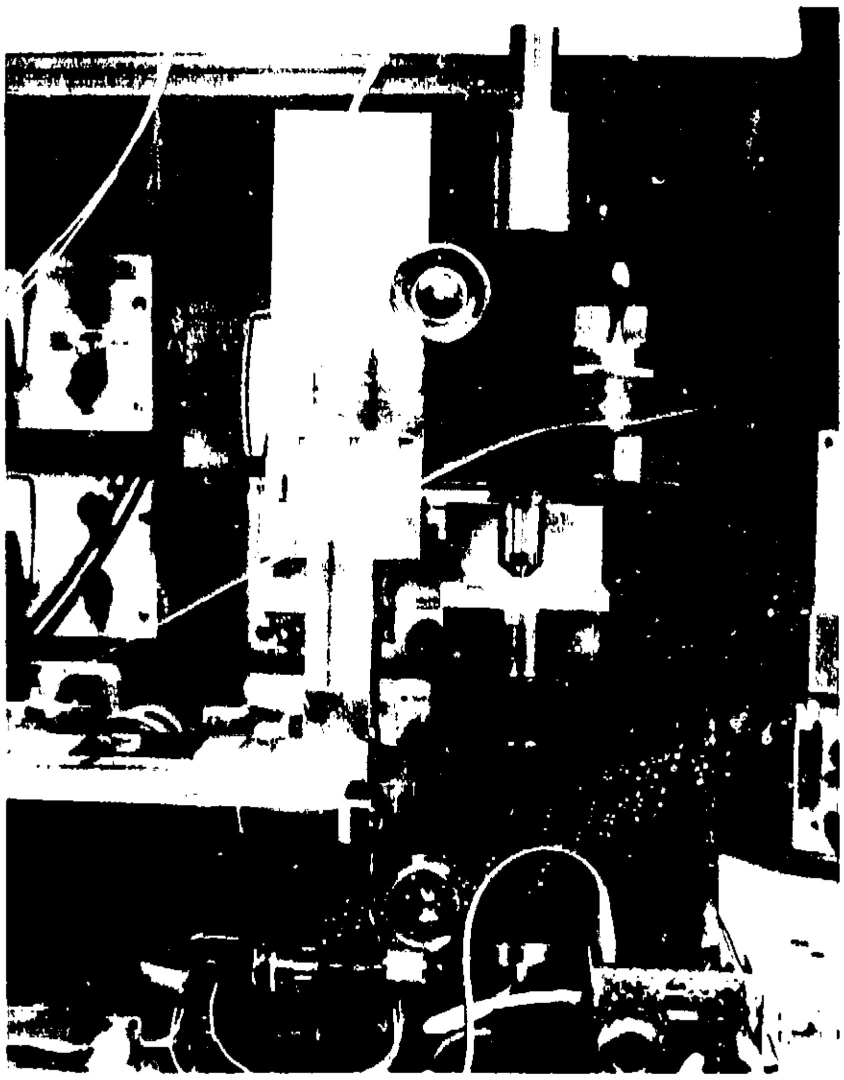


Fig. 6. Optical setup

sensor. The P-N junction used (a GaAs source) is a noncoherent, non-collimated source, but it does have these advantages, particularly applicable to space: it is small, in a solid state, and operates at low voltages (1.5 v). It emits a narrow band of frequencies, around 890 mμ (about 50 mμ wide) and is capable of amplitude modulation at extremely high frequencies (Ref. 1). The modulation capability and the solid-state reliability aspect were prime considerations in using this source. The signal sensor currently is a high-beta, low-signal-level transistor with its TO-5 can removed. It is a silicon transistor and silicon-type junctions are most photo-sensitive to the light at about 925 mμ, hence making an excellent match with the above light source.

By using the amplitude modulating capability of the source and narrow-band amplification of the resultant sensor signal, noise problems are somewhat reduced, especially the low frequency $1/f$ noise of solid-state devices, and the problems associated with dc-coupled amplifiers are avoided. The carrier signal is about 10 kc and, after being amplitude-modulated by transmission through the film, is demodulated either by standard techniques or by a system to be discussed later. The photo current generated in the sensor causes a voltage across a high-impedance parallel network. This voltage is the input signal to a field-effect transistor. Any further signal amplification needed is obtained by conventional techniques. The output signal of the system is, therefore, a 10-kc carrier amplitude-modulated by the various film transmissions in the light path.

The sensor is dc-biased by a small current (150 μa), introduced in the bias-emitter leg (Fig. 7); the 10-kc signal sees a shunt path to ground. V_{CE} is 10 v and the tuned circuit impedance is about 500 kΩ. The optics currently being used are 12.5 mm/0.3 NA for both focusing the spot on the negative and focusing the transmitted light onto the phototransistor. A mimification of 10/1 is obtained with this setup. In the setup shown, the imaged spot on the film is 0.001 in. With an average light source power of 300 mw, an estimated peak signal-to-rms noise of 2/1 has been obtained at 5% transmitted light (Fig. 8). Higher values can be obtained by in-

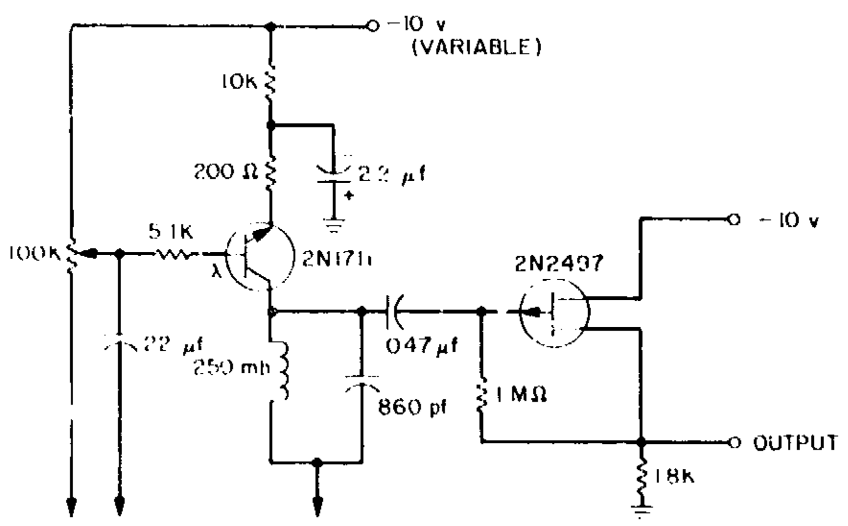


Fig. 7. Photo sensor circuit

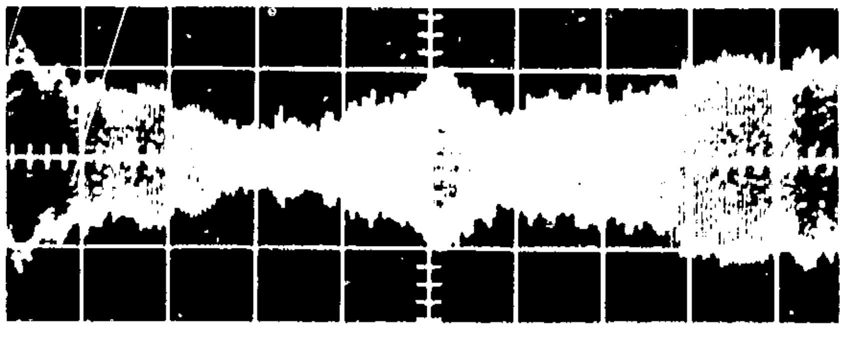


Fig. 8. 5% transmitted light (0.001-in. spot)

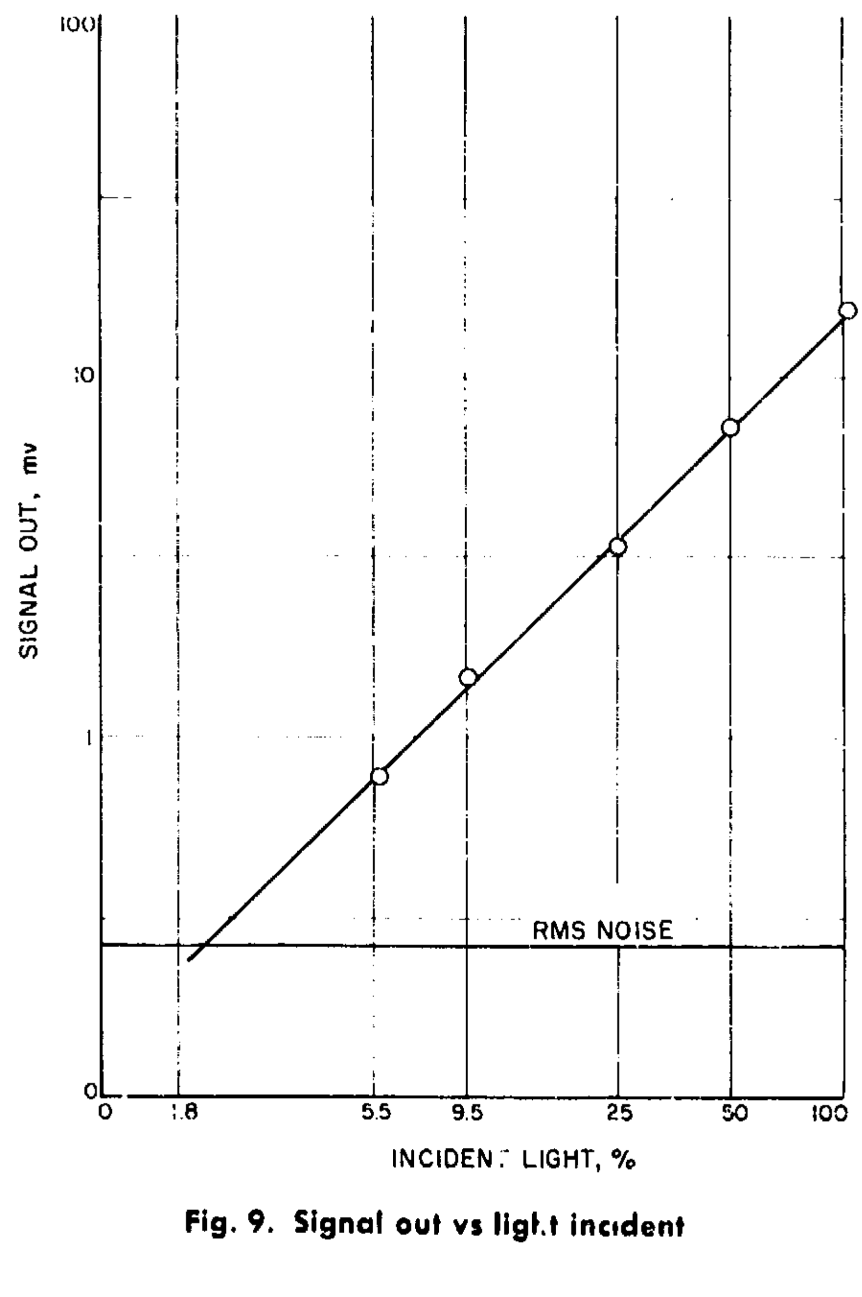


Fig. 9. Signal out vs light incident

creasing the light source power. The source of the noise is primarily in the transistor itself. The same spectrum has been observed with no light source power and is on the order of 10^{-9} amp. This noise is signal-level independent. The curve of signal out vs light incident (Fig. 9) was obtained using Incanel filters corrected for transmission at 890-mμ wavelength.

With this setup an AFTS (black bars on a clear background) was scanned and the following oscilloscope patterns taken of the modulated carrier. Fig. 10 is Group 2 (bar widths go from 0.0049 to 0.0028 in.), Fig. 11 is Group 3 (0.0025 to 0.0014 in.), Fig. 12 is Group 4 (0.0012 to 0.0007 in.).

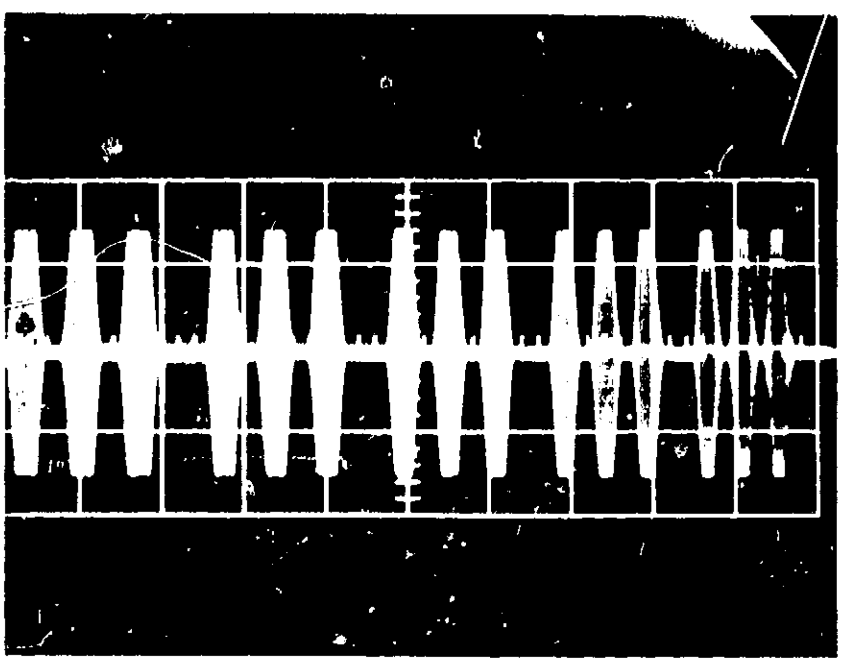


Fig. 10. Group 2 (0.001 spot)

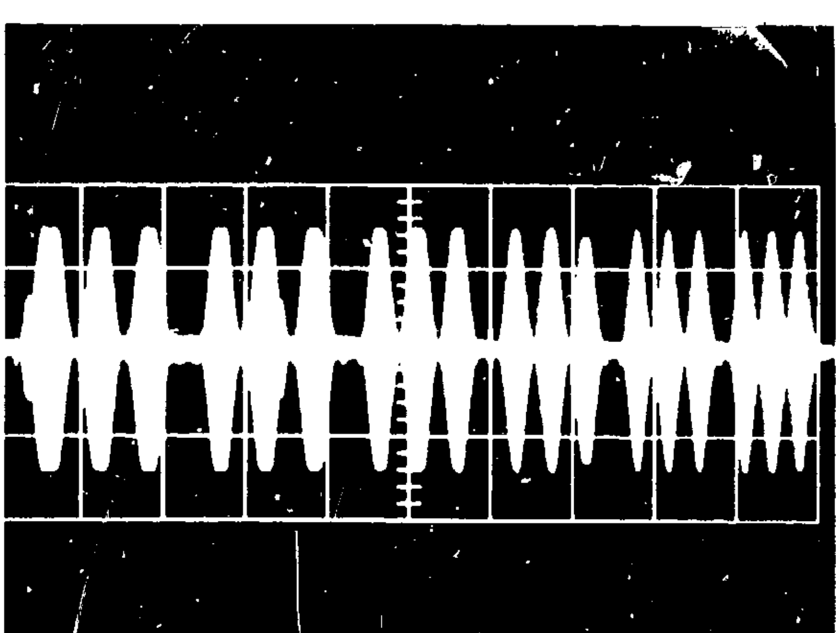


Fig. 11. Group 3 (0.001 spot)

The above figures were coordinated onto a square-wave modulation (Fig. 13). Here the value of the modulation is given in millivolts. The extended line is the rms value of the noise. By extrapolating the curve shown by the dotted line, a peak signal-to-rms noise value of 3/1 occurs at 1700 TV lines/in.

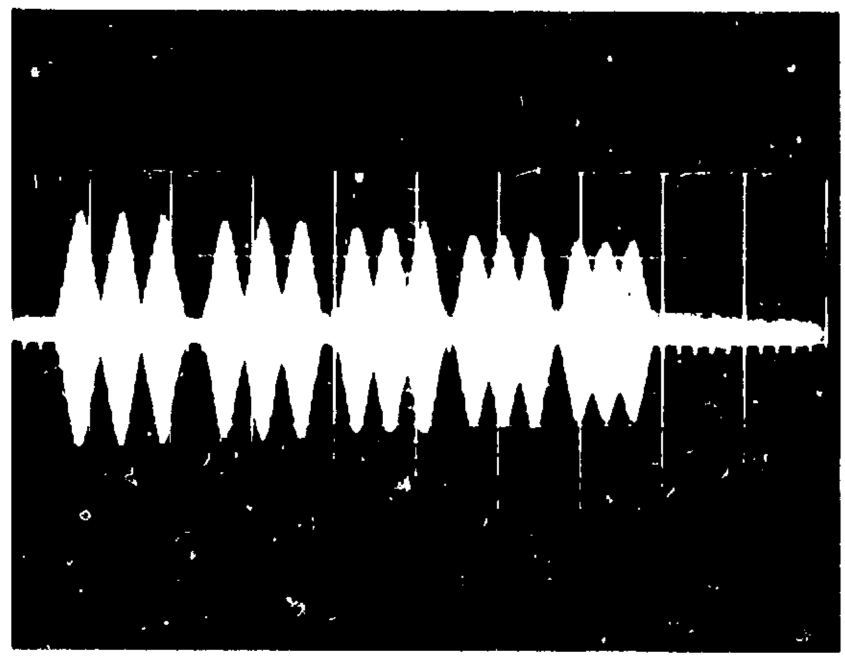


Fig. 12. Group 4 (0.001 spot)

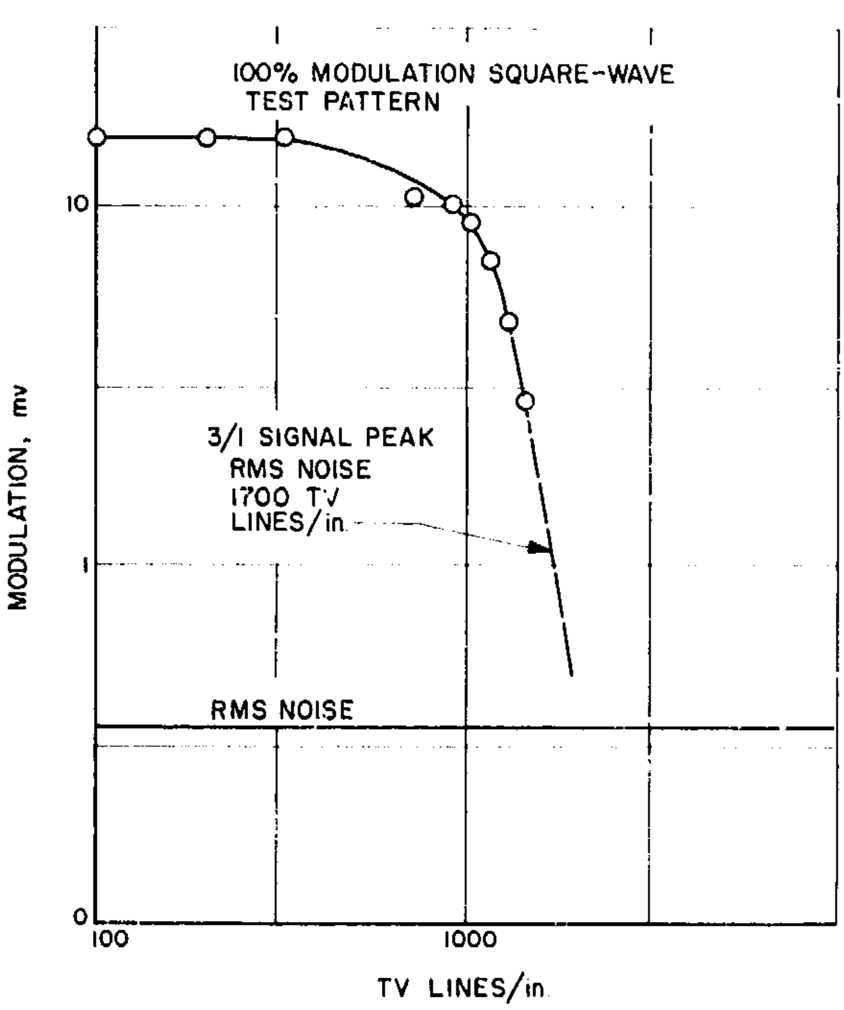


Fig. 13. Square-wave modulation

In order to obtain the necessary frame line scan, the structure shown in Fig. 14 is used. This contains a synchronous motor for line rate and a pulsed stepper motor for frame rate. There is no flyback; the device simply indexes to the next line and reverses. Line spacing can be as small as 0.0008 or integral multiples thereof, and line rate is fixed by the particular gearing at 0.010 in./sec.

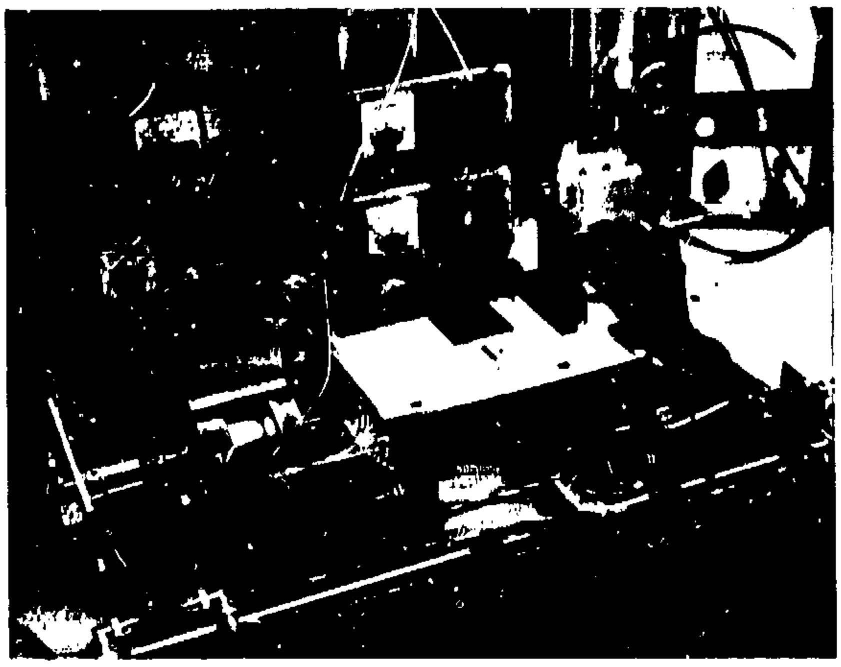


Fig. 14. Overall scan system

There is a potentiometer on each axis for obtaining a position voltage, and these voltages were used to drive a Tektronic 536 horizontal and vertical amplifier. Any area scanned could be scaled up to fill the useful area of the scope. This is far from an ideal setup, but it was one way of being able to present something other than A scope presentations of the data. The 10-kc amplitude-modulated video from the sensor was used to modulate the z-axis of the 536 amplifier. The tube phosphor and the film act as the demodulator for the signal, essentially responding to the envelope alone. An AFTS and a low-contrast Moon scene so reproduced are shown in Figs. 15 and 16.

The system described here is a first try at using the P-N light emitting diodes in this application. The diode used was one of the first series released for general use. Since that time, newer devices acting at higher current levels have been announced. Also the efficiency of the device itself has been increased substantially. It is not unreasonable to expect that a system using more optimally chosen components, or components constructed specifically with the film scanning task in mind, could be built at a nominal cost that would easily give 80 TV lines/mm (0.0005-in. spot size) and a signal-to-noise ratio of better than 100/1. Such a system would use about 2 w for driving the source.

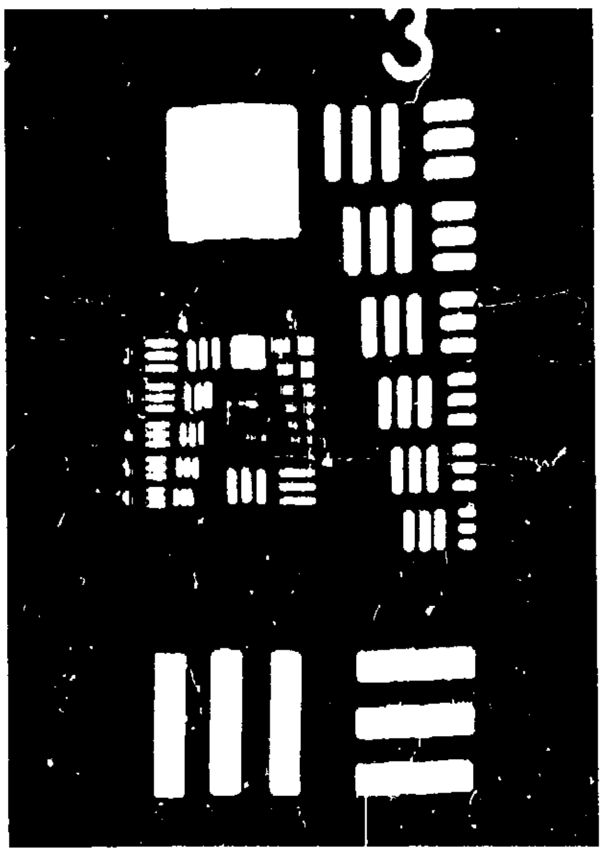


Fig. 15. AFTS reproduced by oscilloscope scan

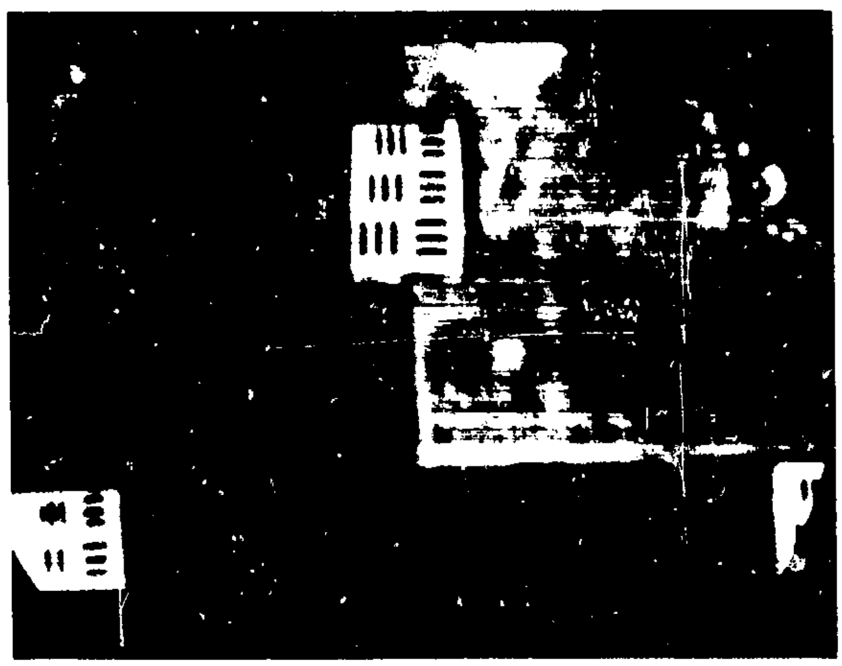


Fig. 16. Low-contrast Moon scene reproduced by oscilloscope scan

C. An Interplanetary Sensitometric Experiment for Radiation Damage Determination

A. W. Spitzak

The problem of determining radiation damage to silver halide films and their developers can be approached in two ways: (1) by simulating, on Earth, the radiation expected and observing the resultant damage, or (2) by constructing a device to perform an engineering experiment to be flown on an interplanetary mission. This mission would be one for which radiation dosages would be representative of those expected on a Martian mission.

A sensitometer device is proposed for use on an interplanetary satellite such as the *Pioneer* series. This device would expose and develop a film sample and measure the density of the exposed and developed areas, noting any increase in base level density due to ambient radiation during the transfer trajectory. The device will duplicate the technique now envisioned for an early reconnaissance mission. Three films will be carried. The film speeds contemplated would be slow, moderately slow, and moderately fast, hence allowing some determination of the potential of these films for space use. To justify the carrying of such an experiment, the device must be as simple as possible without sacrificing the mission objective, and low in weight and power. Two or three experiments might be carried to obtain the time dependence of radiation damage.

1. Experiment Description

A particular film sample will be ¼ × ½ in. At time of use, a ¼ × ¼ in. segment of each sample will receive incident illumination from an internal light source. The other ¼ × ¼ in. sample will receive no incident exposure. The whole film chip will then be processed. To describe what is expected to happen, reference is made to Fig. 17, which contains approximate *D* log *E* curves of the films that might be investigated (Fig. 17a: Kodak plus-X aerecon film; 1/25-sec sunlight; 8-min D-19; 68°F. Fig. 17b: Kodak special fine grain aerial film, grey base, SO-213; 1/10-sec daylight; 8-min D-19; 68°F. Fig. 17c: Kodak special high-definition aerial film, low-speed, SO-243; 1/5-sec daylight; 3½-min D-19; 68°F). These films were exposed to million electron volt values of γ

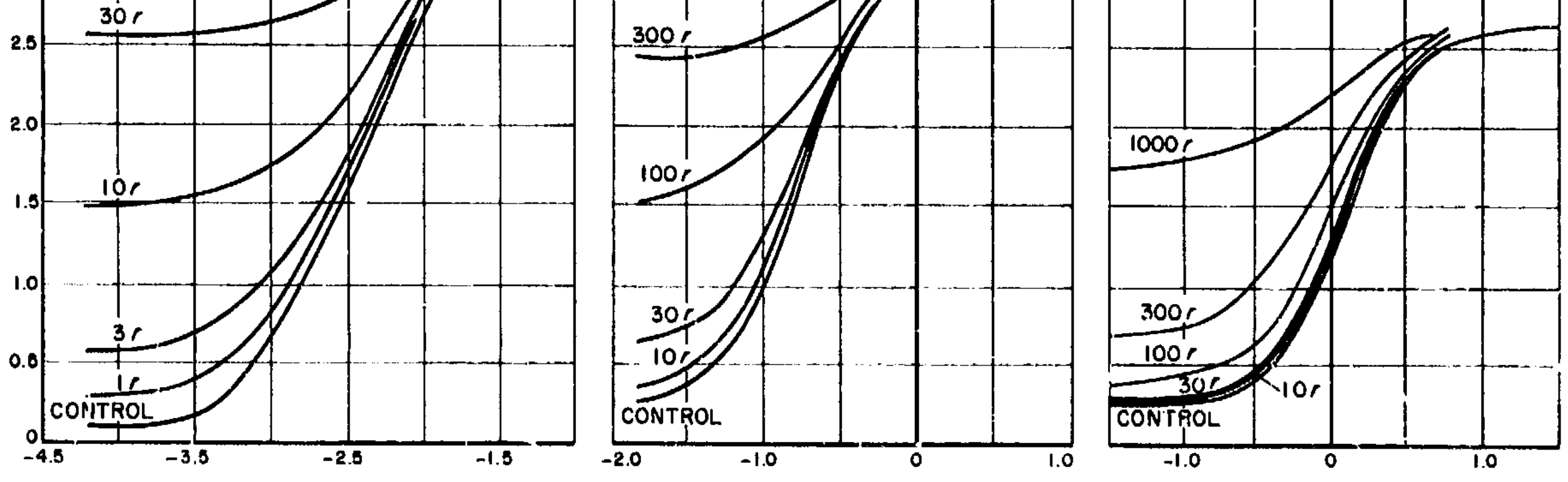


Fig. 17. Radiation damage to films

radiation of various dosages, then their sensitometric curves were plotted in a conventional manner. As may be seen from the curves, the saturation density is independent of the radiation incident upon the film. The purpose of the unexposed and developed cycle is to measure the increase in the fog level of the film resulting from radiation. The purpose of the exposed and developed cycle is to measure the radiation damage to the developer.

The developing section will consist of a monobath-impregnated web chip. Each film will have its own particular monobath chemistry. There are two problems with the developers that must be appreciated and provided for. The first is that, in an oxygen environment, the solutions will oxidize and become useless with time. This can be eliminated by pressurizing the unit with an inert gas (helium, nitrogen). The second problem is the possible drying out of the developer during space travel. This requires maintaining the developer web in a high humidity environment. In this proposal, only the developer environment will be at the high relative humidity; the film chip will be in a dry environment.

After processing and development, a three-position table with ¼-in. steps will be engaged. This table will contain a phototransistor (PT) for each film chip. A diffuse light table which was used previously for exposure will now provide the light for transmission readings. Before the light is turned on, a signal will be obtained from each phototransistor to get a dark current reading. When the light is first turned on, there will be nothing in the path of that light to the phototransistor; hence, a signal will be obtained giving the highest output from the PT. The output of the PT will then be obtained with the film chips in the light paths.

To obtain information at both the low- and high-density ends of the scale will require a dynamic range of 100–1000. A density of 0 corresponds to a 100% transmission, a density of 1 is a 10% transmission, and a density of 2 is a 1% transmission. Naturally, transfer characteristics of the electronics, phototransistors, and lamps must be of known values or capable of being calibrated sufficiently to give the required accuracy.

The actuation members are solenoids. Based on a mechanical power consumption consideration and a 20% efficiency, the electrical power needed is less than a watt. The power consumption will be intermittent using 5 or 6 ½-sec periods over 10 min, indicating the possibility of energy storage, if needed. The two-position solenoid that activates the scan table is outlined in Fig. 18. The drawing shows an exploded view of the sensitometric device. An estimated size of the structure can be obtained from the envelope outline. This is 4 × 2 × 1½ in. A solid block of aluminum of this size would weigh 1.2 lb.

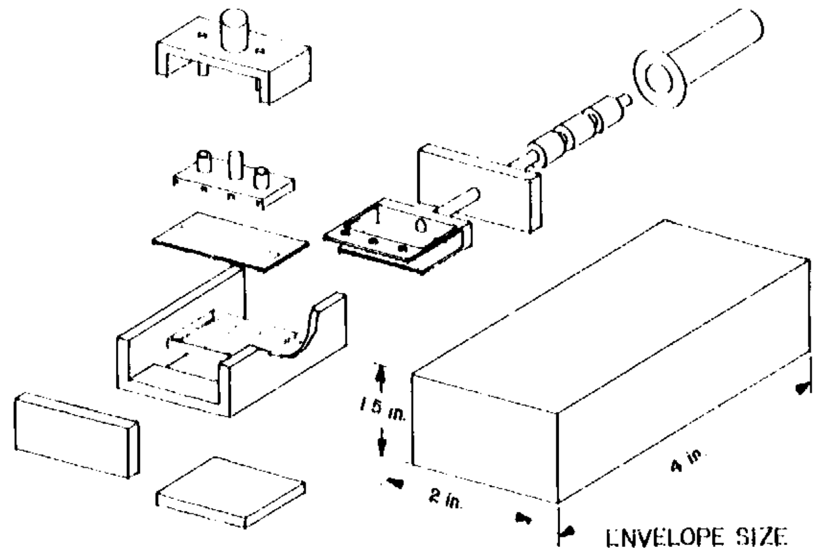


Fig. 18. Exploded view of sensitometric device

Indicative of previous experimentation in the area of radiation damage to film is an article by Aintablian, Gaines, and Lewis describing an experiment where film was flown and recovered (Ref. 2). No exposure or development was handled in space and the dosages experienced were small.

There are several reasons for performing this experiment in the actual environment of intended usage rather than attempting an Earth simulation. First, performing the experiment on the interplanetary trajectory will provide answers as to the acceptability of the sealing mechanism and dehumidified environment of the web processor as well as indicating expected fog levels resulting from radiation. Also, the exposure will occur over an 8-month period. Any Earth-based simulators which made use of high-energy particle accelerators would, of necessity, require very short exposure cycles. Since reciprocity failure (varying exposure time and exposure energy to maintain a constant energy-time product does not result in a constant film density) is a known film phenomenon, the results of an Earth-based simulation would have to be interpreted very carefully. The above statements may best be summarized by saying that performing the experiment under the mission constraints and the mission environment would reduce to a minimum the extrapolation of results as they would apply to a Mars 1969 or 1971 mission.

D. Storage and Space-Camera Data Rates

L. R. Malling

Problems arising with camera data rates and telemetry are discussed. The relationships of storage to video data are outlined. The particular advantages of inserting electronic memory storage in a magnetic-tape system are emphasized. A camera system is described which does not rely on a mechanical shutter or charge-image storage.

1. Space-Camera Data Rates

As the penetration into space becomes deeper and more comprehensive the need to collate the scientific information and control the data rates becomes more pressing. Until the problems with optical communication can be resolved, reliance must continue to be placed on the limited capacity of the present telemetry systems. These require lowered data rates, data compaction, storage, and transmission in non-real-time. These expedients are not always in the best interests of the experiment. Thus, for an impact experiment, a lowered data rate with storage may cause loss of interesting data at the terminal point of the mission. Also, in the instrumentation areas, because of fundamental relationships, $i = dq/dt$ or $e = d\phi/dt$, increases in the time scale force signals into the noise. This is particularly the case for photo-storage TV cameras and for magnetic tape. Some of the disadvantages experienced may be lessened by the insertion of electronic memories or buffers that permit an almost indefinite extension of the time scale without degrading the signal-to-noise ratio (S/N).

The formidable gap that arises between the camera information rate and the telemetry was resolved in the case of *Ranger* by introducing charge-image storage in the camera-tube photoconductor. For planetary-range data this expedient was inadequate; so, for *Mariner* Mars, in addition to extended charge-image storage, electronic memory was added with magnetic tape read-out. In the *Mariner* case, photography formed only part of a scientific instrumentation complex, so that the data requirements of these instruments had also to be recognized. When the resolution and coverage of the photography extends beyond the capability of magnetic tape, as in the lunar orbiter, the video data may be stored on photographic film in the form of a real image. Since the read-out is optical, it is independent of the charge/time relationship, and the time scale can be expanded indefinitely without signal deterioration. Film has a massive storage capability for picture information, ranging to 10^{10} bits/mission.

2. Non-Storage Camera-Tube Advantages

All camera tubes do not use image storage, but for a non-storage tube at normal scanning rates, the S/N is reduced. This may not be an objection when the scene illumination is high, as in sub-solar and terminator photography. Also, there is an advantage in omitting storage in that the output is now independent of the charge/time relationship; the camera information rate can thus be reduced without signal loss. The Farnsworth image dissector is in this category. An expression for the quality of the signal output for the image dissector can be written as $S/N = K(I_s T a/N)^{1/2}$ where I_s is the illumination, T is the frame time, a is the camera aperture size, and N is the number of picture elements. At normal data rates the Farnsworth tube is at a disadvantage in that increasing the resolution elements requires a reduction in the aperture size and, as can be seen, a greatly reduced S/N. However, at an extremely low data rate, where T is greatly increased, this tube may well provide a much better S/N than the storage cameras. A particularly important aspect of this tube is the freedom from shading and target blemishes that so frequently appear in storage surfaces. This is because the photo-image is swept across only a single photo-aperture.

3. Magnetic-Tape Storage

When the telemetry system capacity is extremely limited, as at planetary range, data may be stored on magnetic tape for transmittal at a low data rate at a subsequent time. The tape is a very practical means for collecting data from a plurality of scientific instruments (that include a camera), as demonstrated by *Mariner* Mars. The reliability of a tape transport, its freedom from particle erosion in space, and the ease with which the time scale can be expanded makes its inclusion in a planetary mission particularly appropriate.

Telemetry at planetary range produces an especially wide gap between the camera data rate and the telemetry rate. The solution for *Mariner* included extended charge-image storage in the camera tube. Since the photochemical nature of long-storage photo-surfaces is but imperfectly understood, procurement of slow-scan tubes to meet environmental needs was particularly burdensome. Insofar as greater reliance can be placed on magnetic tape as a storage medium, the emphasis on photo-storage may be reduced or even eliminated.

More effective ways of controlling data rates for scientific instrumentation when using magnetic tape have been investigated (*SPS 37-27*, Vol. IV, p. 101). Electronic memories are inserted into the tape-system mechanization. A two-way electronic buffer through which the tape is both recorded and played back permits indefinite expansion of the time scale while preserving adequate speed and S/N at the tape read-out. In addition, studies have been conducted on high-density recordings up to 10,000 bits/lineal in. (*SPS 37-27*, Vol. IV, p. 97). Modification of camera operation to take advantage of these improved techniques may permit, in selected cases, quite simple shutterless-camera operation without the need of a slow-scan mode.

4. Charge-Image Storage Problems

A distinction must be made in a slow-scan tube between photo-storage, which is photon integration of the light flux, and charge-image storage, which is electro-capacitance storage that takes place on the same photosensitive surface. This dual role has been found to be particularly effective and simply mechanized for space applications with the vidicon. The reduced information rate is secured by a slow-scan read-out of the charge image. The data rate may be expressed for a frame time T as $H = zN/T$ bits/sec where z is the quantization of the picture elements N. A picture element is defined by the pitch Δx, where the element area $\Delta A = (\Delta x)^2$.

Slow-scan vidicon operation is made possible by shuttering the light flux. The signal read-out is no longer sequential and the frame rate is a compromise between the desired information rate, the discharge characteristics of the photo-surface, and the need to introduce special scan modes in order to stabilize the charge image. The dq/dt loss of the slow scan is compensated insofar as possible by maximizing the integration of the light flux and optimizing circuit parameters (*SPS 37-19*, Vol. VI, p. 79). The slow-scan signal loss is not compensated by the reduced bandwidth requirement, since the amplifier noise component $i_n = (kTB)^{1/2}$. Further, because of stray capacitance, noise pick-up or transistorization, it is not feasible to raise the load impedance to a value commensurate with a greatly reduced scanning rate.

The impact of a slow scan on vidicon performance can be appreciated where, in the case of Nimbus, with a frame time of 200 sec, the S/N for the relatively high Earth illumination is only 20 db. The expedient of beam pulsing was only partially successful, since the bandwidth requirement increased from 1.6 to 25 kc with a corresponding noise increase.

The exposure time is eventually limited by resolution degradation caused by image motion. The motion is that produced by the relative movement of the spacecraft to the observed scene. If the information rate is to be reduced with a shutterless vidicon, image motion may become the limiting factor. Whether a shutter is used or not, a maximum exposure is generally desired to preserve a high S/N.

5. Shutterless-Camera Without Charge-Image Storage

If the photographic objectives can be realized by increasing the number of frames and reducing the information/frame, the possibility arises of dispensing both with charge-image storage and the use of a mechanical shutter. Some reorientation of the system operating modes is required to permit this. A Mars fly-by mission may be used to identify these modes.

The assumed photographic objective is a swath across the planet photographed at a specified resolution for a transit of the spacecraft across the planetary diameter. During the transit, time must be allowed for data to accumulate from associated scientific instrumentation, i.e., continuous data storage by TV is not permitted. This requires that a number of pictures be taken, separated in time but with overlapping areas.

For a strip 400 km wide, at an optical resolution of 5 km/line pair, 3600 lines must be scanned during a transit time of 20 min. Each line will contain 200 elements with a picture element size $\Delta A = 2 \times 2$ km for a Kell factor of 0.8.

Since $H = zN/T$, the information rate can be reduced by reducing N and maximizing T. This is illustrated in Fig. 19 for a reduced N, where the format has been

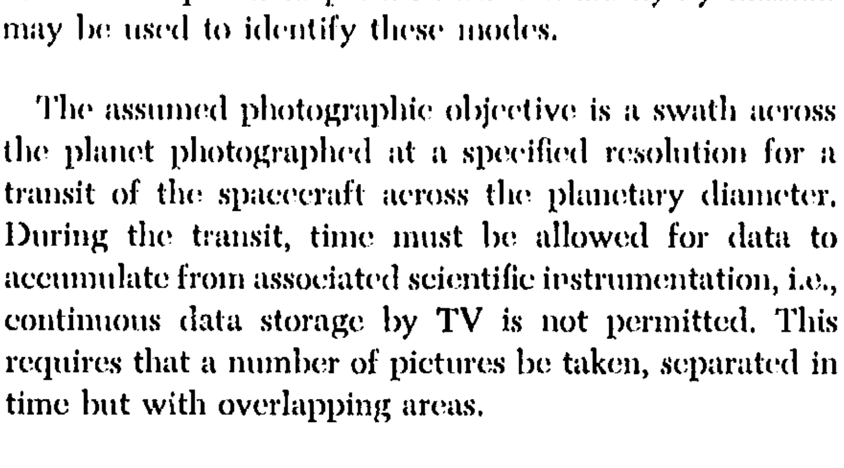


Fig. 19. Frame format for shutterless scanning on camera tube face

reduced from 1.0 × 1.0 to 1.0 × 0.1, encompassing only 20 lines/frame while retaining 200 elements/line. Two hundred frames with 10% overlap provide the desired coverage at the specified resolution. The frame time T, which is now also the exposure time, may be maximized to the point of image smear.

There are two aspects to image smear: spacecraft motion and photo-image decay. Spacecraft motion is concerned with: (1) transit of the field-of-view across the planetary surface, and (2) the spacecraft as a stable platform. For a 20-min transit at 25,000-km range, the image motion is 3 lines/sec as produced by a transit sweep of 5.5 km/sec. For an image movement of the order of $\Delta x/2$, the exposure must be limited to 0.2 sec. This gives a frame rate of 5/sec. The roll component is less than 2 min arc/min. For an exposure of 0.2 sec, this only produces 0.03 Δx image displacement. Thus, for the camera, the spacecraft can be considered as a stable platform.

Photo-image decay is expressed as a percentage remaining-residual over the frames scanned. A typical figure for a broadcast vidicon is 0.4 of the previous image remaining over a single scan time. For succeeding images the decay is $(0.4)^n$, where n is the number of intervening frames. This may well be improved by the insertion of special wiping modes, selection and preparation of the photo surface, or other expedients currently used with slow-scan vidicons.

The effect of image smear on resolution is illustrated in Fig. 20, which takes a *Ranger VIII* photograph and superimposes a 0.5 contrast image at selected image shifts. Fig. 20a shows the original photograph, Fig. 20b demonstrates 0.5Δx shift, and Fig. 20c demonstrates 2Δx shift. While the 2Δx shift is readily noticed, the 0.5Δx shift is scarcely perceptible. Δx is the picture element width for a 200-line scan.

6. Shutterless-Camera System

For the 1200-sec transit and the 200 pictures, the picture taking rate is one every 6 sec. Science data will be accumulated in the intervening periods. A simplified mechanization is shown in Fig. 21. Logical control selects only one of the many pictures taken during the 6-sec period for encoding. The highest signal frequency generated

$$f_c = \frac{N(\Delta A)}{2T} = \frac{20 \times 200}{2 \times 0.2} = 10 \text{ kc}$$

The encoder sampling time $\Delta t = 1/2 f_c = 50\ \mu$sec. For 6-bit encoding, a data bit occupies 8 μsec. This is well

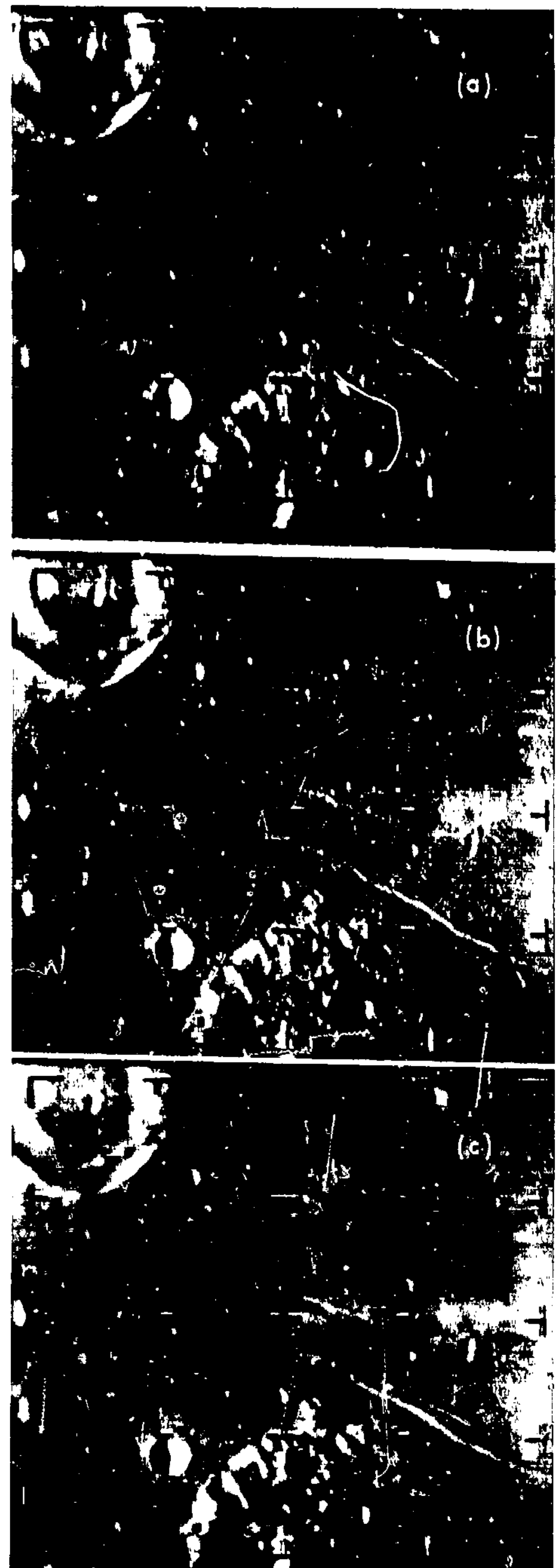


Fig. 20. Loss of resolution produced by image smears (Ranger VIII photographs)

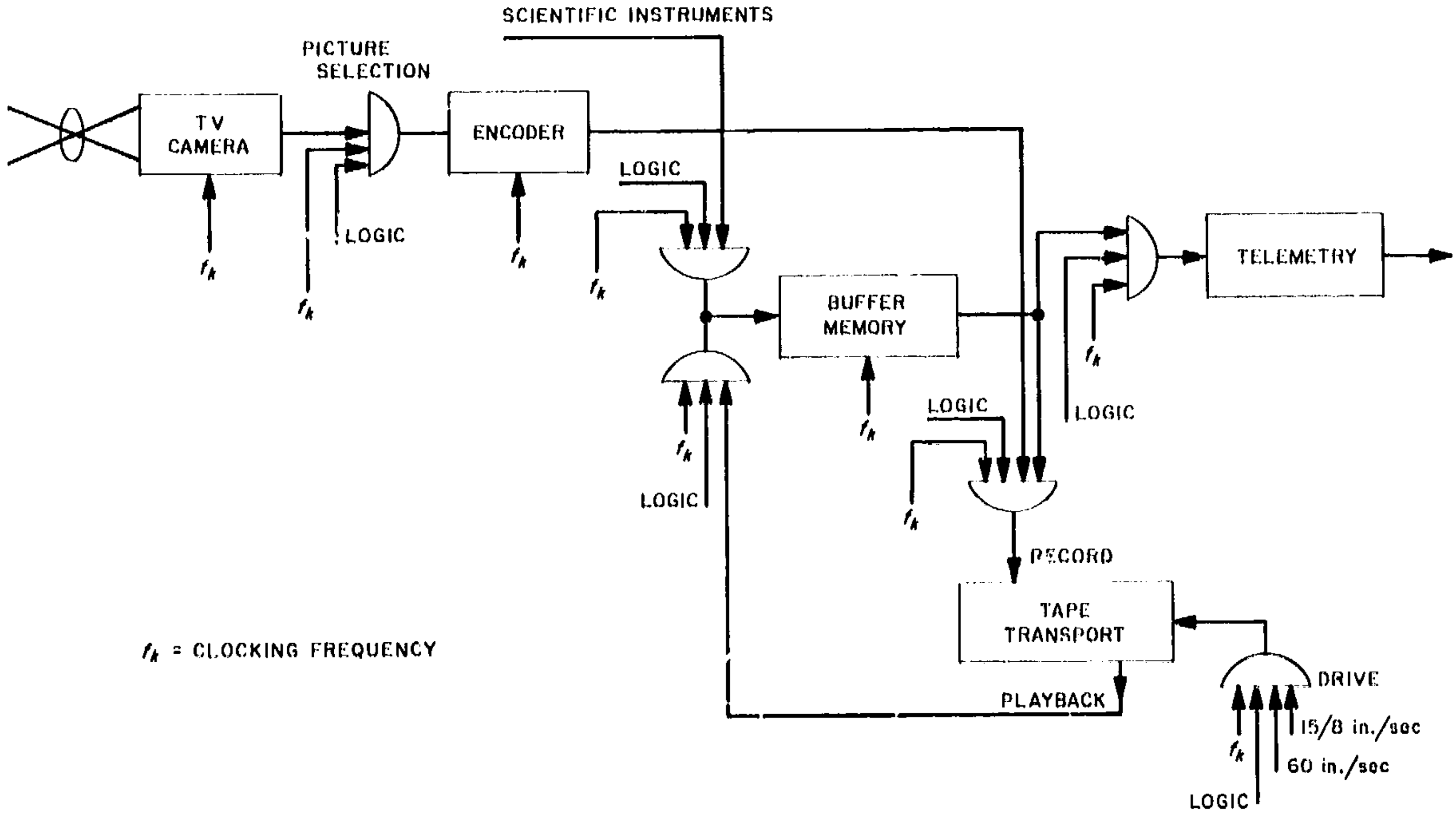


Fig. 21. Simplified block diagram of shutterless TV camera system

within the capacity of the *Mariner* digital encoder. The data rate generated by the camera H = 120K bits/sec for 1200 bits/line or 24K bits/frame. If space cannot be allocated for electronic memory of 24K bit capacity, two tape recording speeds are required with direct recording of the video information, as shown in Fig. 21. Typical recording speeds are 60 in./sec for the camera and $1\frac{7}{8}$ in./sec for the science data. Acceleration and deceleration of the tape consume a substantial part of the available tape for a nominal 1 + 1 sec allowance with a total tape time of 2.2 sec/frame. Constancy of speed is not a must during video recording, since the interposition of the electronic buffer smoothes all data proceeding to telemetry. At 60 in./sec, the packing density is 2K bits/in. The tape length for the video alone, assuming 4 active data tracks, would be 550 ft.

With completion of the data acquisition phase, the tape transport is unloaded at 15/8 in./sec. Video data is unloaded at line intervals only into the buffer for subsequent relaying to telemetry, nominally operating at $8\frac{1}{3}$ bits/sec. The telemetry time lost, taken by buffer reloading, will be 0.32 sec, less than 3 bits for the 1200-bit line. The telemetry time taken to unload the buffer will be 2.4 min/line or 4.8 min/frame. The more significant parameters have been included in Table 1.

Table 1. Significant parameters for shutterless camera system

Transit time	1200 sec
Number of lines (total)	3600
Number of frames	200
Picture format	1.0 × 0.1
Lines/frame	20
Resolution (optical)	5 km
Frame rate	5/sec
Number of picture elements/frame	4000
Picture element (dimension)	1.8 × 1.8 km
Encoded frame rate	1 frame/6 sec
Line scanning rate	10 msec/line
f_c	10 kc
Encoder rate	120K bits/sec
Frame capacity	24K bits
Tape speed (record)	60 and 15/8 in./sec
Tape speed (playback)	15/8 in./sec
Tape tracks	4
Acceleration and deceleration time	1 + 1 sec
Total tape time/frame	2.2 sec
Tape packing density	2K bits/in.
Active tape/frame	12 in./frame
Required tape length/frame	11 ft
Line time (playback to buffer)	0.32 sec
Line time (buffer to telemetry)	2.4 min
Lineal length for video requirement	2200 ft
Tape length	550 ft

References

1. Rediker, R. H., *Solid State Design*, p. 19, August 1963.

2. Aintablian, A. A., Gaines, E. E., and Lewis, J. C., "Effects of Penetrating Radiations in Space on Photographic Materials and Other Detectors," *Photographic Science and Engineering*, Vol. 7, No. 3, May–June 1963.

XVIII. Space Instrument Systems

A. Derivation of Timing and Control Signals for the Ultraviolet Spectrometer Data Automation System

M. Perlman

1. Introduction

Present *Mariner*-class telecommunication systems utilize a single-channel, serial, binary data system. Scientific data are time-division multiplexed onto this channel. This article is confined to the problem of deriving timing and control signals for multiplexing data from several scientific instruments for transmission in real-time. A real-time system is herewith defined as a system in which the instruments are sampled immediately prior to transmission of the sample and/or in which data storage is limited to one sample (typically 10 bits). The *Mariner C* (1964 Mars mission) spacecraft included a real-time system (*SPS 37-24*, Vols. I, II, and III). The work reported here represents an effort to improve the apparatus and techniques for use on future missions.

2. Data Frame Format

A data frame will consist of 420 serialized binary digits. Two blocks of 140 bits each are reserved for non-real-time (NRT) science data. Two blocks of 70 bits each are reserved for real-time (RT) science and engineering data. A possible data format follows:

140 NRT	70 RT	140 NRT	70 RT

Included in the 420 bits will be identification and control words.

3. Period and Subperiod Generation

a. Division of frame period. In order to provide a maximum of flexibility and make the data automation system as mission-independent as possible, it is proposed that all possible subperiods (subframe signals) be made available to the user. The full period is 420 bits. A one-to-one correspondence exists between the subperiods and the divisors of 420.

The "Fundamental Theorem of Arithmetic" states that every positive integer $N > 1$ may be uniquely represented (except for order) as the product of one or more primes. A prime is defined as an integer p, if $p > 1$ and p has no positive divisors except 1 and p.

Thus,

$$420 = 2^2 \cdot 3 \cdot 5 \cdot 7 \tag{1}$$

The divisors of 420 may be expressed in terms of its prime (power) factors.

Let $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, and $p_4 = 7$. Consider the expression

$$(1 + p_1 + p_1^2)(1 + p_2)(1 + p_3)(1 + p_4) \tag{2}$$

Multiplying out Expression (2) results in a polynomial of 24 terms. Each term is a divisor of 420. For example, $p_1p_2p_3$ equals 30, a divisor of 420. One of the 24 divisors, namely 1, is trivial. The 23 non-trivial divisors are tabulated in Tables 1 and 2. A prime or prime power is a factor of the divisor if a 1 appears under its column heading. Note that either 2^2 or 2, but never both, may be a factor of a given divisor.

Table 1. Divisors of 420

2^2	2	7	5	3	Non-trivial divisors of 420
0	0	0	0	1	3
0	0	0	1	0	5
0	0	0	1	1	15
0	0	1	0	0	7
0	0	1	0	1	21
0	0	1	1	0	35
0	0	1	1	1	105
0	1	0	0	0	2
0	1	0	0	1	6
0	1	0	1	0	10
0	1	0	1	1	30
0	1	1	0	0	14
0	1	1	0	1	42
0	1	1	1	0	70
0	1	1	1	1	210
1	0	0	0	0	4
1	0	0	0	1	12
1	0	0	1	0	20
1	0	0	1	1	60
1	0	1	0	0	28
1	0	1	0	1	84
1	0	1	1	0	140
1	0	1	1	1	420

Table 2. Divisors of 420 in ascending order

2^2	7	3	Non-trivial divisors of 420
2	10	28	84
3	12	30	105
4	14	35	140
5	15	42	210
6	20	60	420
7	21	70	

b. Implementation of period and subperiod timing and control signals.

Techniques of scaling. Scalers synthesized from logical memory and decisional elements will be used to derive period and subperiod timing and control signals. A scaler may be defined as a digital subunit which delivers one output for every τ inputs. Scales of 2^2, 2, 3, 5, and 7, with the input of each scaler coming from the system timing source of t bits per sec, are all that are necessary to derive any scale (i.e., divisor) of 420. By logically *anding* the outputs of various scalers in accordance with Tables 1 and 2, signals for every 2, 3, 4, ·····, 140, 210, and 420 system timing (source) inputs may be made available.

Logical elements. The basic decisional element is the *nand* (*not-and*) gate with an expandable input and provisions for common collector operation with other *nand* gates. A logical amplifier will be used for increasing the output drive capability of a basic decisional element or a memory element. An *anding* structure is incorporated at the amplifier's input. Up to three inputs may be logically *anded*. One of the important uses of the logical amplifier is to provide drive to the trigger circuits of the memory elements. The memory element is a set-reset flip-flop made of cross-coupled *nand* gates with independent reactive triggering for the set and reset enable inputs.

With few exceptions the memory elements are to be operated synchronously. That is, the states of the memory elements will be permitted to change only during the clock (i.e., timing source) pulse interval. The logical symbols for the *nand* gate, logical amplifier, and set-reset flip-flop appear in Figs. 1a, b, and c.

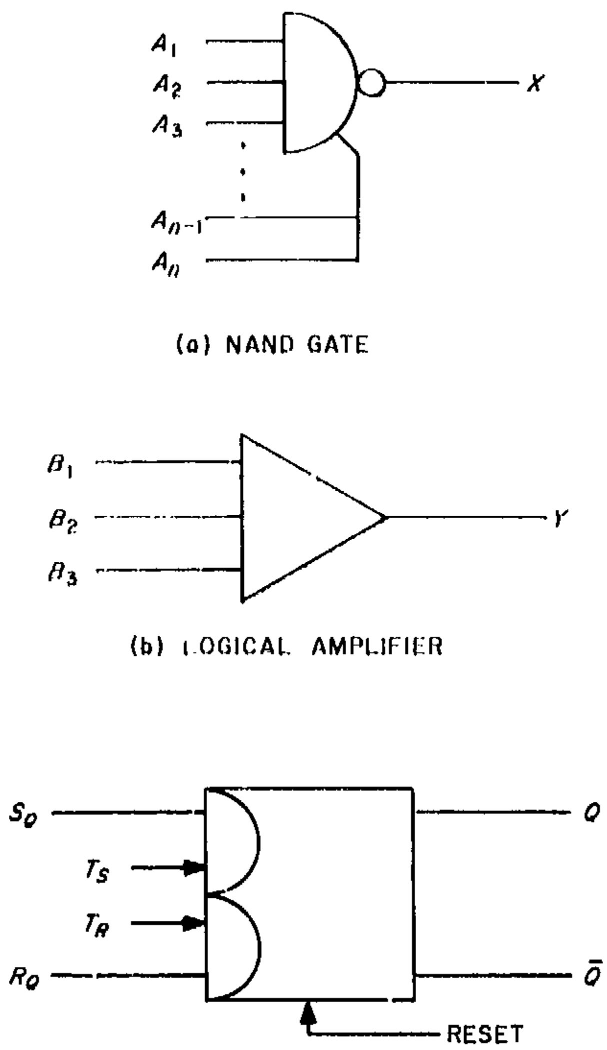


Fig. 1. Logical symbols

The Boolean expressions which characterize the output of the *nand* gate and the logical amplifier as a function of their inputs are:

$$X = \overline{A_1 A_2 \cdots\cdots A_{n-1} A_n} \tag{3}$$

$$Y = B_1 B_2 B_3 \tag{4}$$

Expressions (3) and (4) are valid for the correspondence between voltage levels and logic (state) values of

$$0 \text{ v} \longleftrightarrow \text{Logic } 0$$

$$+ \text{ v} \longleftrightarrow \text{Logic } 1$$

For the *nand* Function (3), an output of Logic 0 results when all inputs are at Logic 1. For all other input combinations, the output of the *nand* gate is a Logic 1. An unused input leg assumes a voltage level which corresponds to a Logic 1.

With the logical amplifier, the *and* operation is realized [Expression (4)]. An output of Logic 1 results when all inputs are at Logic 1. For all other input combinations, the output of the logic amplifier is at Logic 0. An unused input leg assumes a voltage level which corresponds to a Logic 1.

The set-reset flip-flop which serves as a memory element has two types of inputs. These are voltage levels S_Q and R_Q, and voltage transitions T_S and T_R. S_Q and R_Q are the set enable and reset enable inputs. The set gate and reset gate of the flip-flop are preconditioned by the levels represented by the Boolean variables S_Q and R_Q. The assertion output Q is also defined as the present state of the flip-flop. The next state Q' is a function of S_Q, R_Q, and Q, which are present at the time a negative-going voltage transition appears at T_S and T_R. Note that, with few exceptions, T_S and T_R are electrically connected. (Independent triggering will be used in rare instances of asynchronous sequential logic applications.) The trigger $T = T_S = T_R$ is implied in the characteristic equation of the flip-flop, which is derived as follows: Table 3, the truth table, characterizes the behavior of the memory element. The next output (i.e., next state) Q', defined in terms of the present state Q and present enable inputs, is assumed in approximately 100 nanoseconds after a negative-going voltage transition appears at T.

Table 3. Truth table for set-reset flip-flop

R_Q	S_Q	Q	Q'
0	0	0	$\cancel{0}$
0	0	1	$\cancel{0}$
0	1	0	0
0	1	1	0
1	0	0	1
1	0	1	1
1	1	0	0
1	1	1	1

A Logic 0 represents the voltage level which enables S_Q for setting or R_Q for resetting. Note that if both S_Q and R_Q are enabled (i.e., $S_Q = R_Q = \cancel{0}$), the next state Q' is indeterminate and is represented by 0. The next state or output Q' in minimized disjunctive form appears in Equation (5), which is a Boolean difference equation.

$$Q' = \bar{S}_Q + R_Q Q \tag{5}$$

also,

$$\bar{R}_Q \bar{S}_Q = 0 \tag{6}$$

Equation (6) represents a constraint. No attempt should be made to set and reset the set-reset flip-flop simultaneously, since the next state is indeterminate. In the synthesis of a sequential circuit, the response (i.e., next state Q') requirements of each memory element are first determined. The response requirement of a memory element is shown in generalized form in the Boolean difference Equation (7).

$$Q' = g_1 Q + g_2 \bar{Q} \tag{7}$$

Q is the present state and g_1 and g_2 are Boolean functions of the present state values of other memory elements and external inputs. Thus, from

$$\bar{S}_Q + R_Q Q = g_1 Q + g_2 \bar{Q} \tag{8}$$

where $\bar{R}_Q \bar{S}_Q = 0$, S_Q and R_Q must be determined. In generalized form S_Q and R_Q are Boolean functions of g_1, g_2, and Q. Note that additive and multiplicative cancellation (subtraction and division) are undefined in Boolean algebra. Therefore, S_Q and R_Q cannot be determined by methods employed in ordinary algebra. The inverse of the characteristic Equation (5) is needed where S_Q and R_Q input expressions are derived from the required response Q'. This may be done from truth table characteristics of Equations (8) and (5), given in Table 4.

The R_Q and S_Q entries can be verified in Table 3. For example, in Row 1, the present state Q is 0 and remains 0 for the next state Q'. This corresponds to row entries

Table 4. Truth table derivation of R_Q and S_Q inputs

g_1	g_2	Q	$Q' = g_1Q + g_2\bar{Q}$	R_Q	S_Q
0	0	0	0	c_1	1
0	0	1	0	0	1
0	1	0	1	1	0
0	1	1	0	0	1
1	0	0	0	c_2	1
1	0	1	1	1	c_3
1	1	0	1	1	0
1	1	1	1	1	c_4

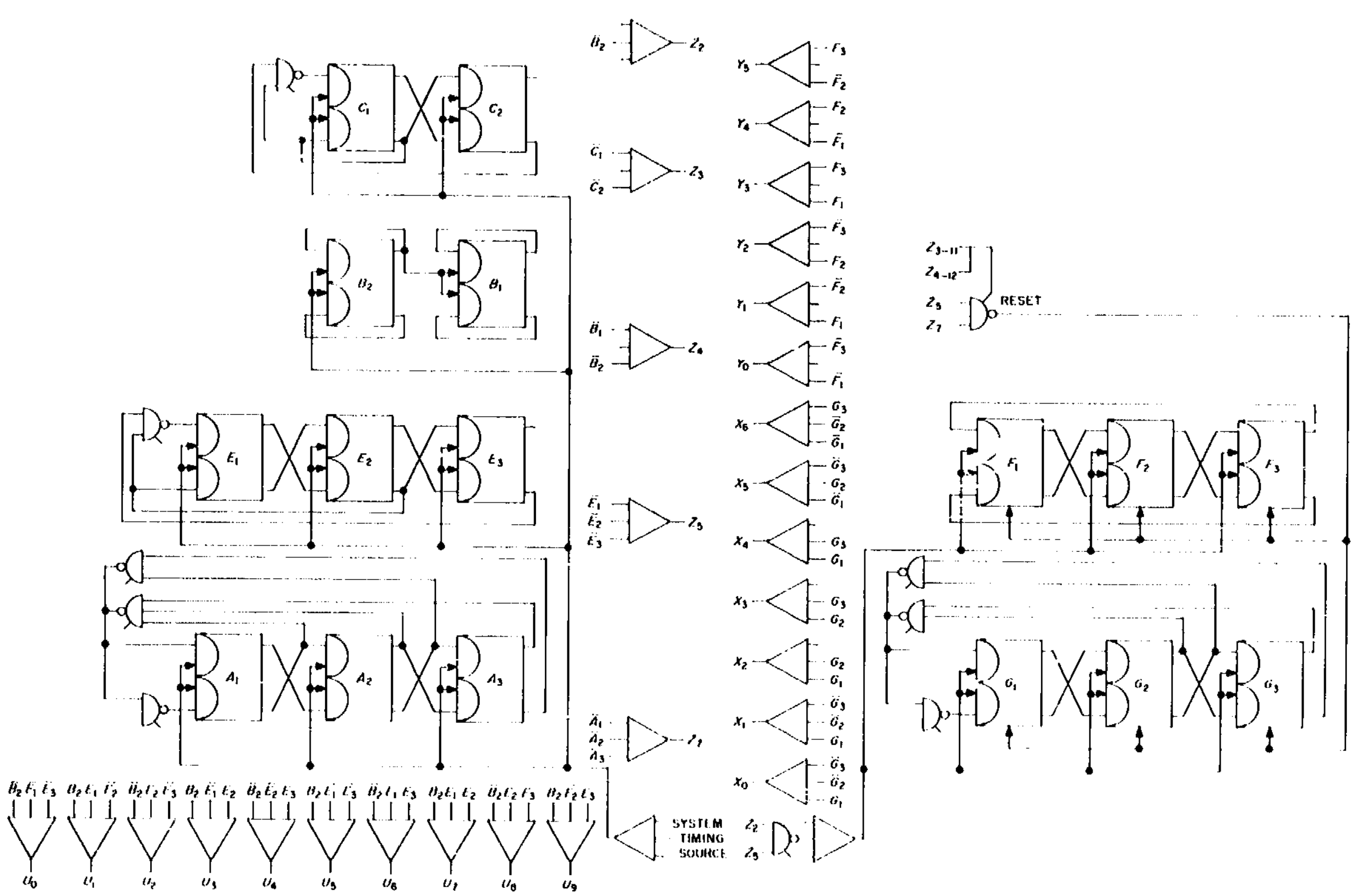


Fig. 2. Science timing and control signals

3 and 7 in Table 3, which show that when Q is at 0, Q' will be 0 if R_Q is 0 or 1 and S_Q is at 1. Returning to Row 1 of Table 4, the entry under R_Q is c_1, a Boolean constant which may be assigned the logic value 0 or 1 arbitrarily, and the entry under S_Q is 1. Table 4 may be simplified and merged to four rows, since there are only four distinct combinations of present to next-state transitions (Table 5). Note that variable g_1 or g_2 is eliminated in each disjunctive term.

Table 5. Simplification of Table 4

	Q	Q'	R_Q	S_Q
$\bar{g}_2\bar{Q}$	0	0	a	1
$g_2\bar{Q}$	0	1	1	0
$\bar{g}_1 Q$	1	0	0	1
$g_1 Q$	1	1	1	b

The simplified table yields minimum disjunctive forms of generalized expressions for R_Q and S_Q.

$$R_Q = a\,\bar{g}_2\bar{Q} + g_2\bar{Q} + g_1 Q \tag{9}$$

$$S_Q = \bar{g}_2\bar{Q} + \bar{g}_1 Q + b g_1 Q \tag{10}$$

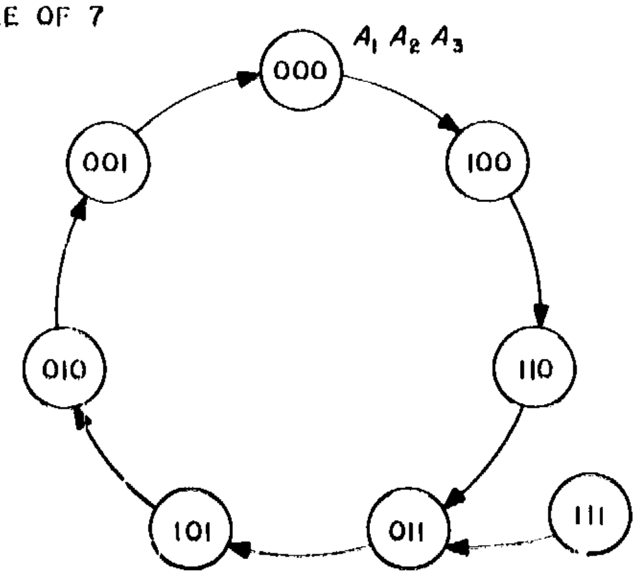
The Boolean constants a and b are assigned the logic value (0 or 1) which yields the simplest expression in terms of implementation. These assignments are dependent upon g_1 and g_2 (i.e., the required response of the memory element).

Implementation of scalers. Economy of logical elements was one important consideration used in the synthesis of the scalers. Synchronous operation of the memory elements was assumed unless asynchronous operation yielded a significant reduction in logical elements. One memory element is equivalent to approximately three decisional elements, excluding packaging considerations.

Internal state assignments which yielded a minimum of combinational logic were chosen. The results of applying Equations (9) and (10) for the chosen state assignments are shown for each scaler. Successive internal states appear in a state diagram. Note that unused (combinations of) states are shown. In each case, they are succeeded by a state which is part of the main scaling cycle. This assures that a malfunction due to a transient error is self-correcting. With the exception of the scaler which generated scales of 2 and 4, each scaler is basically a shift register with logic feedback.

The implementation of the scalers is shown in Fig. 2.

SCALE OF 7



$$A'_1 = \bar{A}_1A_2A_3 + \bar{A}_2\bar{A}_3$$
$$R_{A_1} = \bar{A}_1A_2A_3 + \bar{A}_2\bar{A}_3$$
$$S_{A_1} = \bar{R}_{A_1}$$

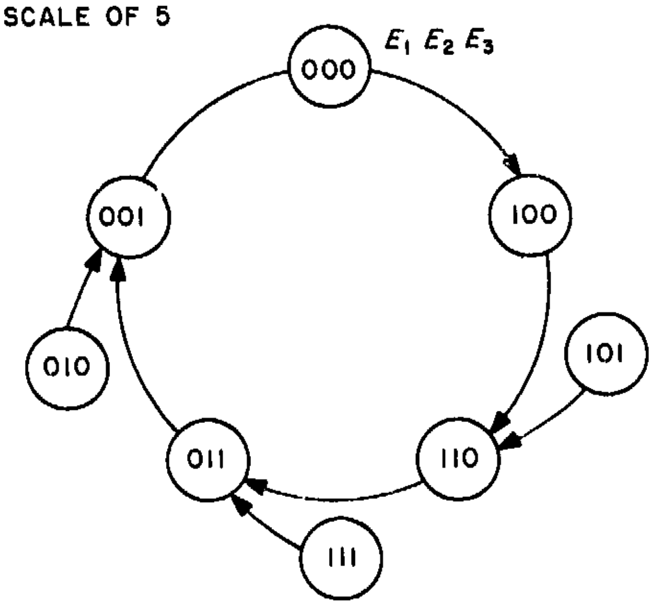
$$A'_2 = A_1$$
$$R_{A_2} = A_1$$
$$S_{A_2} = \bar{A}_1$$

$$A'_3 = A_2$$
$$R_{A_3} = A_2$$
$$S_{A_3} = \bar{A}_2$$

$$Z_7\ (\text{output}) = \bar{A}_1\bar{A}_2\bar{A}_3$$

Cost······3 flip-flops
3 *nand* gates
1 power amplifier

SCALE OF 5



$$E'_1 = E_1\bar{E}_2 + \bar{E}_2\bar{E}_3$$
$$R_{E_1} = \bar{E}_2$$
$$S_{E_1} = \overline{\bar{E}_2\bar{E}_3}$$

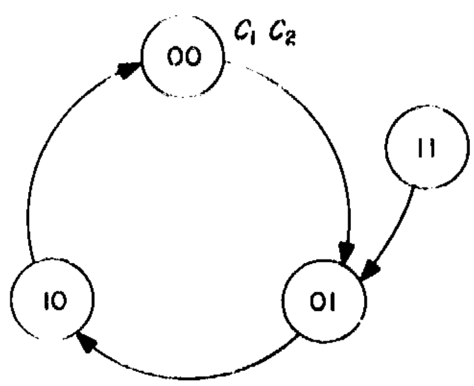
$$E'_2 = E_1$$
$$R_{E_2} = E_1$$
$$S_{E_2} = \bar{E}_1$$

$$E'_3 = E_2$$
$$R_{E_3} = E_2$$
$$S_{E_3} = \bar{E}_2$$

$$Z_5 = \bar{E}_1\bar{E}_2\bar{E}_3$$

Cost · · · · · 3 flip-flops
1 *nand* gate
1 power amplifier

SCALE OF 3



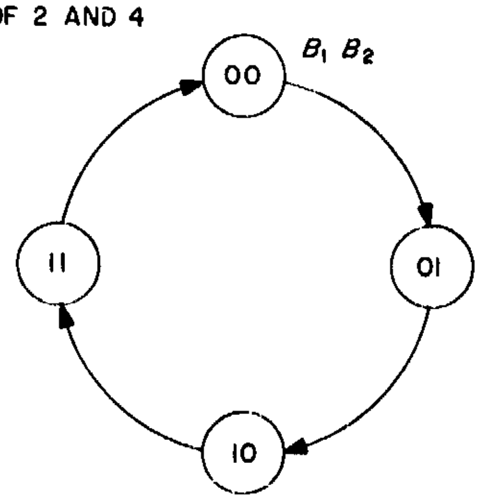
$$C'_1 = \bar{C}_1\bar{C}_2$$
$$R_{C_1} = \bar{C}_1$$
$$S_{C_1} = \overline{\bar{C}_1\bar{C}_2}$$

$$C'_2 = C_1$$
$$R_{C_2} = C_1$$
$$S_{C_2} = \bar{C}_1$$

$$Z_3 = \bar{C}_1\bar{C}_2$$

Cost · · · · · 2 flip-flops
1 *nand* gate
1 power amplifier

SCALE OF 2 AND 4



$$B'_1 = T_{B_1}\,\bar{B}_1 + \bar{T}_{B_1}\,B_1$$

$$T_{B_1} = B_2$$

$$R_{B_1} = \bar{B}_1$$

$$S_{B_1} = B_1$$

$$B'_2 = T_{B_2}\,\bar{B}_2 + \bar{T}_{B_2}\,B_2$$

$T_{B_2} = P$ (system timing, i.e., clock pulse)

$$Z_2 = \bar{B}_2$$

$$Z_4 = \bar{B}_1\bar{B}_2$$

Note that T is a Logic 1 only during a negative-going voltage transition.

Cost · · · · · 2 flip-flops
2 power amplifiers

4. Word Counting and Bit Stripping

a. Word counting. The data frame may be divided into a whole number of words of any fixed length (shown in Table 3). For example, there are 42 10-bit words in the 420-bit data frame. A Modulo 42 counter may be employed to keep track of the data words, 10 bits in length, within the frame. Since a scale of 2 and a scale

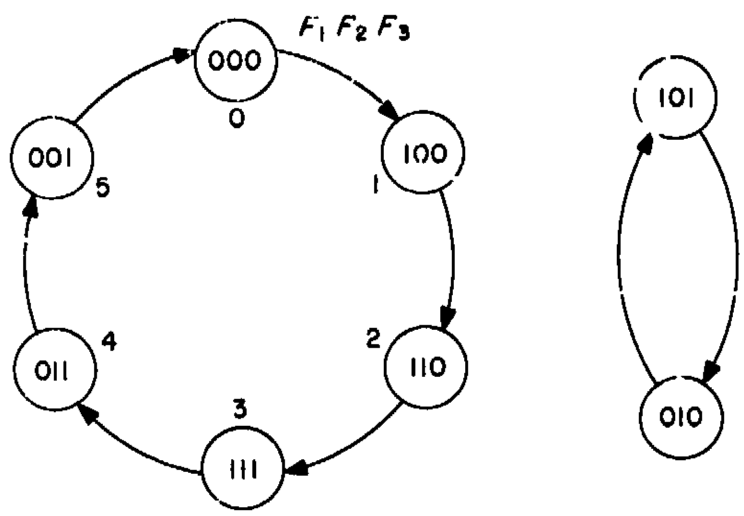
Table 6. Data word versus X_iY_j

i	j	Data word
0	0	1
1	1	2
2	2	3
3	3	4
4	4	5
5	5	6
6	0	7
0	1	8
1	2	9
2	3	10
.	.	..
.	.	..
.	.	..
.	.	..
4	3	40
5	4	41
6	5	42

of 5 are available, *anded* outputs of these scales could be used as trigger inputs to a Modulo 6 and Modulo 7 counter, respectively. A matrix decoder which accepts the 6 distinguishable counts out of the Modulo 6 counter and the 7 distinguishable counts out of the Modulo 7 counter can furnish 42 ordered pairs of outputs (since 6 and 7 are relatively prime). Each ordered pair corresponds to one and only one 10-bit data word within each augmented one count for every 10 bits emanating from the system timing source.

Implementation of counters. The same criteria used in the synthesis of scalers were applied to the synthesis of counters. Successive combinations of states do not represent successive binary numbers. The counting codes adopted are non-weighted. This results from state assignments chosen to minimize the combinational logic.

MODULO 6 COUNTER



$F_1' = \bar{F}_3$ $\quad R_{F_1} = \bar{F}_3$ $\quad S_{F_1} = F_3$

$F_2' = F_1$ $\quad R_{F_1} = F_1$ $\quad S_{F_1} = \bar{F}_1$

$F_3' = F_2$ $\quad R_{F_3} = F_2$ $\quad S_{F_3} = \bar{F}_2$

$Y_0 = \bar{F}_1\bar{F}_3$ $\qquad Y_3 = \bar{F}_1\bar{F}_3$

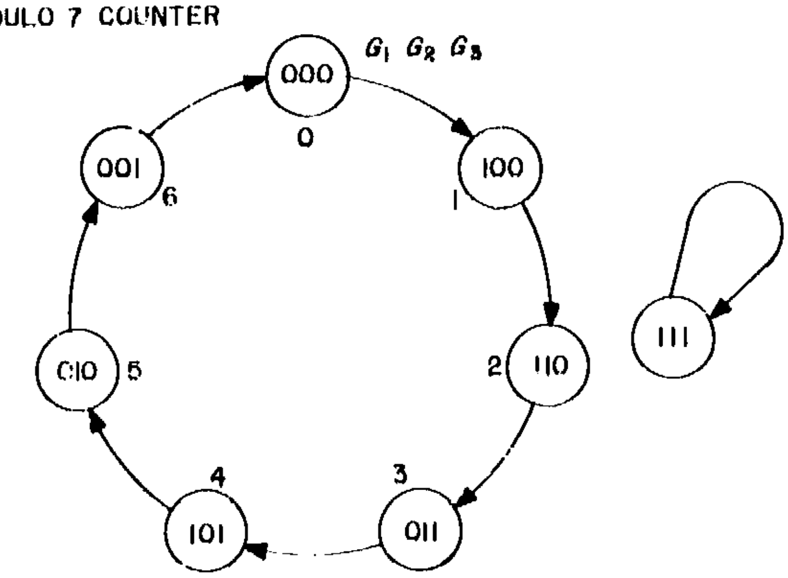
$Y_1 = F_1\bar{F}_2$ $\qquad Y_4 = \bar{F}_1F_2$

$Y_2 = F_2\bar{F}_3$ $\qquad Y_5 = \bar{F}_2F_3$

Note that protection against the counter falling into the minor cycle is provided by a reset signal which is generated during the 420th bit of each frame. All flip-flops in the Modulo 6 and 7 counters receive the reset signal (see Fig. 2).

Cost · · · · · 3 flip-flops
6 power amplifiers

MODULO 7 COUNTER



$G_1' = G_2G_3 + \bar{G}_2\bar{G}_3$

$R_{G_1} = G_2G_3 + \bar{G}_2\bar{G}_3$

$S_{G_1} = \bar{R}_{G_1}$

$G_2' = G_1$ $\quad R_{G_2} = G_1$ $\quad S_{G_2} = \bar{G}_1$

$G_3' = G_2$ $\quad R_{G_3} = G_2$ $\quad S_{G_2} = \bar{G}_2$

$X_0 = \bar{G}_1\bar{G}_2\bar{G}_3$ $\qquad X_4 = G_1G_3$

$X_1 = G_1\bar{G}_2\bar{G}_3$ $\qquad X_5 = \bar{G}_1G_2\bar{G}_3$

$X_2 = G_1G_2$ $\qquad X_6 = \bar{G}_1\bar{G}_2G_3$

$X_3 = G_2G_3$

Cost · · · · · 3 flip-flops
3 *nand* gates
7 power amplifiers

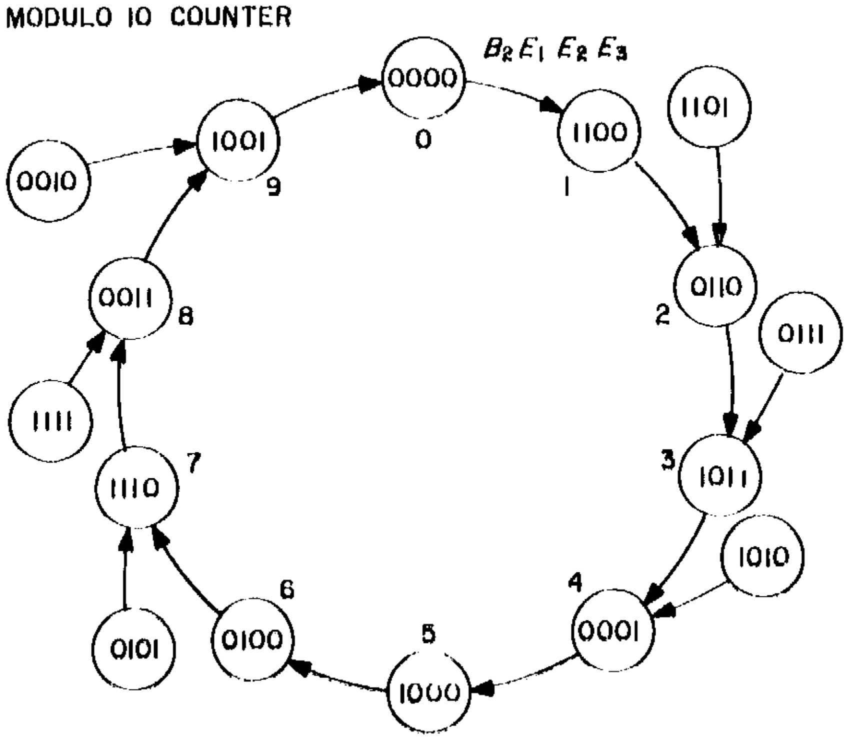
As indicated in the first paragraph, Part 4a, each *anded* ordered pair X_iY_j (for $0 \leq i \leq 6$, $0 \leq j \leq 5$) corresponds to one and only one 10-bit data word. A partial table of this correspondence appears in Table 6.

Thus, Data Word 40 will be generated during the time the *anded* output X_4Y_3 is at the voltage level representing a Logic 1.

b. Bit stripping. Each bit within a 10-bit data word can be identified with a Modulo 10 counter operating at the system's timing (source) rate. The 10 distinct state combinations that the scale of 2 combined with the scale of 5 pass through may be decoded to provide bit time information.

Implementation of the bit stripper. A state diagram of the internal state changes of the Scalers 2 and 5 is shown. Minimized output functions are also given.

MODULO 10 COUNTER



$$U_0 = \bar{B}_2\bar{E}_1\bar{E}_3 \qquad U_5 = B_2\bar{E}_1\bar{E}_3$$

$$U_1 = B_2E_1\bar{E}_2 \qquad U_6 = \bar{B}_2E_1E_3$$

$$U_2 = \bar{B}_2E_2\bar{E}_3 \qquad U_7 = B_2E_1E_2$$

$$U_3 = B_2\bar{E}_1E_2 \qquad U_8 = \bar{B}_2E_2E_3$$

$$U_4 = \bar{B}_2\bar{E}_2E_3 \qquad U_9 = B_2\bar{E}_2E_3$$

Cost · · · · · 10 power amplifiers

By *anding* U_k $(0 \leq k \leq 9)$ with X_iY_j, the *k*th bit of a particular word can be extracted.

The means for counting 42 10-bit data words and stripping a particular bit (or a combination of particular bits) is shown in Fig. 2. Only the inputs to the 6×7 matrix appear.

XIX. Chemistry

A. Soil Studies—Microflora of Desert Regions. III. Microorganisms in Valley of 10,000 Smokes Desert

F. A. Morelli, R. E. Cameron, and G. B. Blank

1. Introduction

Extraterrestrial life detection experiments are being designed and instrumented for detecting and characterizing microorganisms on Mars. These experiments are designed to sample a Martian soil or soil-like surface materials (Refs. 1 and 2). In preparation for these extraterrestrial investigations, studies on terrestrial desert soils may provide useful background information for designing life detection experiments on Mars.

In desert regions, soils have developed primarily under the environmental influences of an arid climate, a relatively long geological time period, a rugged barren topography, and parent materials which have been little modified in soil formation except in terms of smaller size particles. Martian environmental factors for soil formation will undoubtedly be even more severe. Available water and vegetation are the two primary factors in development and formation of terrestrial soils. These factors have a minor role in genesis of most terrestrial desert soils and may be nonexistent on Mars.

The Valley of 10,000 Smokes desert in Katmai National Monument, Alaska is a unique terrestrial desert (Figs. 1 and 2). This volcanic area has sufficient moisture, but is devoid of macrovegetation, except at the periphery. This desert was formed in 1912 following a series of earthquakes, the subsequent eruption of Mt. Katmai, the formation of a vent now existing as a volcanic plug, Novarupta, and the opening of hundreds of fumeroles and fissures, the latter issuing hundreds of tons of pumice, ash, and tuff-like materials from fresh rhyolitic and andesitic magmas. Finally, an area of more than 42 mi^2 was covered to a depth of more than 700 ft in some places by some 7 mi^3 of volcanic debris (Fig. 2). Four years later, this area was investigated by the first of successive National Geographic Society Expeditions (Ref. 3), although it was not until 1932 that any microbiological studies were reported for the area (Ref. 4). No microbiological information has been reported since then, except for a survey report in 1959 of a few species of mosses, lichens, fungi, and algae (Ref. 5).

Fig. 1. Valley of 10,000 Smokes (sample location, topography, and vegetation distribution)

Fig. 2. Valley of 10,000 Smokes (view from the northwest)

2. Soil Sample Sites

In August 1963, 12 samples of soil and soil-like materials were obtained by aseptic technique from the vicinity of the Valley for microbiological determinations. Additional samples were obtained for determining pertinent physical and chemical properties of the samples. A complete record of physical, chemical, and microbiological analysis of these samples has now been achieved.[1]

[1] Cameron, R. E., Mamikunian, G., Morelli, F. A., and Blank, G.B., "The Record of Katmai Soils. 1. Soil Characteristics" (to be published).

A summary of the characteristics of four of these soils is given in Tables 1–5. Location of collection sites for all 12 of these samples is shown in Fig. 1. For the four samples represented in this article, the sample sites were as follows: (1) Site 115, a barren, pebbly surface 1 in. deep, approximately 200 yd northeast of footbridge over Ukak River, northwest of Three Forks at west end of Valley (Fig. 3); (2) Site 118, a barren surface 1 in. deep,

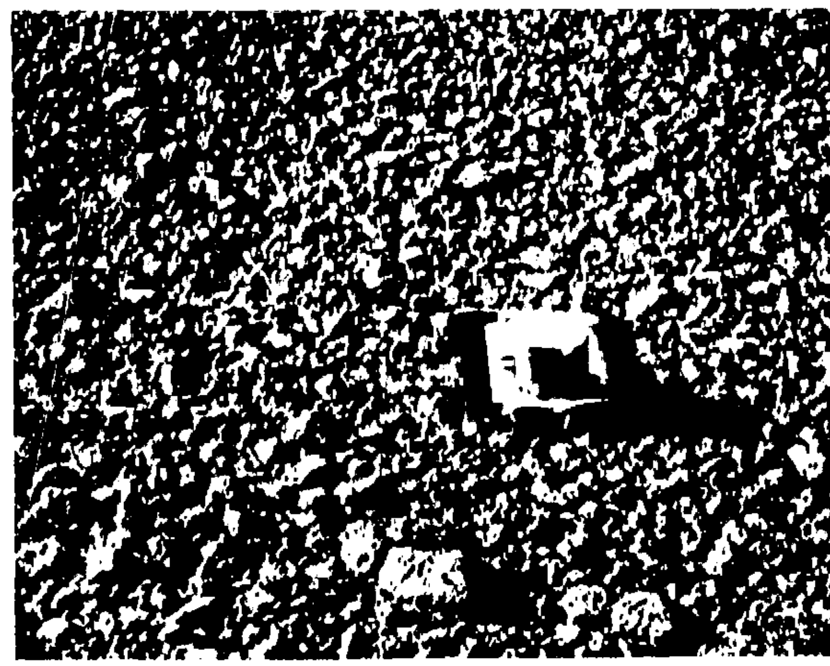


Fig. 3. Sample Site 115 (lichen-encrusted coarse pumice)

Table 1. Valley of 10,000 Smokes soil physical determinations

Soil	Color (air-dry)	Textural class	Porosity, %	Moisture determinations, % H_2O			
				In situ	Field capacity	Wilting coefficient	Hygroscopic coefficient
115	Light-gray	Sandy loam	69.8	13.2	6.9	6.8	0.4
118	Red	Loam	66.7	16.6	14.0	5.5	1.0
119	White	Loam	67.5	20.1	15.3	13.0	1.0
122	Pale brown	Silt loam	66.4	31.7	11.2	8.1	2.3

Table 2. Valley of 10,000 Smokes soil physico-chemical determinations

Soil	pH[a]	Eh, + mv[a]	Electrical conductivity, $\kappa \times 10^6$ mhos/cm at 25°C[a]	Osmotic pressure, atm	Cation exchange capacity, meq/100 g of soil
115	4.6	500	27	0.004	2.5
118	4.7	550	17	0.006	2.0
119	4.9	580	20	0.002	2.5
122	4.4	520	52	0.008	9.0

[a]Saturated soil paste

Table 3. Valley of 10,000 Smokes soil chemical determinations

Soil	Total organic matter, %	Carbon-nitrogen values							
		Total carbon, %	Organic carbon, %	Carbonate carbon, %	Total nitrogen, %	Organic nitrogen, %	Nitrate nitrogen, %	Ammonia nitrogen, %	C:N total
115	0.73	0.11	0.15	0.00	0.020	0.013	0.0001	<0.0001	5.6
118	0.11	0.07	0.09	0.00	0.020	0.007	0.0001	<0.0001	3.4
119	0.08	0.06	0.05	0.00	0.020	0.010	0.0001	<0.0001	3.2
122	15.9	5.9	5.5	0.00	0.190	0.190	0.0009	0.0060	3.1

Table 4. Valley of 10,000 Smokes soil microbiological determinations (general microflora)

Soil	Microorganisms per gram of soil				
	Aerobes	Microaerophiles	Anaerobes	Fungi	Algae
115	4700	1000	12,500	1900	100,000
118	190	1000	0	0	10,000
119	0	10	5	0	0
122	5200	10,000	15	15	100
Range for 12 soils	0 to 12,600	10 to 100,000	0 to 12,500	0 to 1900	0 to 100,000
Media	Trypticase soy agar	Fluid thioglycollate medium	Trypticase soy agar + CO_2	Rose bengal agar	Pochon solution

Table 5. Valley of 10,000 Smokes soil microbiological determinations (specialized microflora)

Soil	Microorganisms per gram of soil				
	Heterotrophic iron oxidizers	Autotrophic iron oxidizers	Sulfur oxidizers	Thiosulfate oxidizers	Iron-sulfur metabolizers
115	1000	0	10	10	1000
118	1000	0	100	10	1000
119	10	0	0	10	100
122	1000	1000	10	1000	1000
Range for 12 soils	10 to 1000	0 to 1000	0 to 1000	0 to 1000	10 to 1000
Media	Iron ammonium-citrate nitrate solution	Leathen's solution	Starkey's solution	Beijerinck's thiosulfate solution	Iron-sulfur solution

bank of extinct, reddish upraised fumerolic incrustment by the side of a shallow ravine, approximately ¼ mi south of base of Mt. Jule (Fig. 4); (3) Site 119, a barren surface 1 in. deep, approximately 1¾ mi west of Site 113, edge of extinct, whitish, fumerolic incrustation on steep west-facing bank of tributary of Knife Creek (Fig. 5); and (4) Site 122, a mossy surface 1 in. deep, patch of open soil between *Equisetum* and various tall grasses, approximately 250 yd north of new observation cabin at northeast end of Valley (Fig. 6).

Fig. 4. Sample Site 118 (bank of extinct fumerole)

Fig. 5. Sample Site 119 (fumerolic material from bank of Knife Creek)

3. Soil Analyses

Physical determinations showed a range of colorful soils of various textures and high porosities. All soils contained sufficient available *in situ* moisture contents. Soil reaction values were all below pH 5.0 and typical of volcanic deserts. These soils were well oxidized, contained very little soluble salts and had very low cation exchange capacities, even for a relatively organic soil (Sample 122). Nearly all of the carbon and nitrogen, obtained by different methods, was in the organic form in the barren soils. C:N values were very narrow and typical for arid desert soils[2] (Ref. 6).

Microbiological determinations were performed in the laboratory on 10-g portions of soil, and incubations were carried out at 25°C in order to determine basic groups of microflora (Table 4). When low numbers of microflora were obtained for some of the samples, e.g., Samples 118 and 119, additional media were utilized that would be favorable for culture of specialized microorganisms, which, on the basis of physical and chemical data and the environmental situation, might be present in the Valley. Determinations for these organisms are shown in Table 5, and include the iron and sulfur bacteria. Ranges for numbers of microorganisms in the 12 soils are also shown in Tables 4 and 5. For all of the five major groups of organisms, including aerobes, microaerophiles, anaerobes, fungi, and algae, at least one or more of these soils contained none of these groups, except the microaerophiles, which were present in all soils. The microaerophiles and algae were the most abundant groups present. For 11 of the 12 soils, all of these counts are lower than those obtained for *typical* desert soils.[2]

The algae count is highest for Soil 115, which shows desert-type algal-lichen-moss incrustations. The microaerophilic count is highest for soils either outside the Valley or at the periphery of the Valley. Proximity of vegetation to sample site has a definite influence on total numbers of these basic groups of microorganisms (Table 6). Proximity of vegetation also has an apparent influence on the predominant group of microorganisms present. Soils 115 through 121 were obtained in the Valley proper and show a predominance of algae in 5 out

Table 6. Valley of 10,000 Smokes soil (abundance of soil microflora with proximity of vascular vegetation)

Soil	Depth of soil	Number of microflora per gram of soil	Major group present	Distance from vegetation, ft (approx.)
115	Surface 1 in.	12×10^5	10^5 algae	25
116	↓	10,030	10^4 algae	5,280
117	↓	1,250	10^3 algae	165
118	↓	12,980	10^4 algae	1,320
119	↓	20	Iron-sulfur bacteria	7,920
120	↓	100	10^2 algae	7,920
121	↓	35	Iron-sulfur bacteria	7,920
122	↓	17,550	10^4 microaerophiles	0.5
123	1–6 in.	19,210	12.6×10^4 aerobes	0.5
124	Surface 1 in.	13,030	10^4 microaerophiles	25
125	Surface 1 in.	24,770	10^4 algae 10^4 microaerophiles	25
126	1–6 in.	102,075	10^5 microaerophiles	25

[2]Cameron, R. E., Morelli, F. A., and Blank, G. B., *Desert Soil Characteristics. I. Preliminary Studies*, Technical Report, Jet Propulsion Laboratory (to be published).

of 7 soils. Soils 112 through 126 were obtained either outside the Valley or at the periphery of the Valley and showed an abundance of microaerophiles. The only soil which had a predominance of aerobes, 123, was a subsurface sample of ejected ash underlying newly formed soil. A transition soil, 125, consisted of older soil washed over ejected ash; it showed a predominance of both algae and microaerophiles, and under this soil the highest number of microorganisms, the microaerophiles, was obtained for all of the 12 soils. No actinomycetes were noted in any of these soils, as was indicated previously (Ref. 4), although they are noticeable inhabitants of typical desert soils. Fungus counts were exceptionally low or nil, especially inside the Valley. Soils obtained in the middle of the Valley contained the fewest organisms, and these were predominantly iron and sulfur organisms, but heterotrophes rather than chemosynthetic autotrophes. A complex iron-sulfur medium containing salts plus desert soil extract was synthesized and found to be most favorable for growth of these organisms (see Footnote 1, p. 198)

Typical microorganisms obtained from these soils are shown in photomicrographs (Figs. 7 through 10). The algae are represented by only a few species, and restricted primarily by the low pH. A coccoid green alga, *Protococcus grevillei* (Fig. 7), is the most abundant alga in these soils and is also the predominant green alga in typical desert soils. A representative fungus is the ascomycete *Penicillium* sp. (Fig. 8), also a dominant organism in desert soils. Bacteria in these soils include primarily species of (1) nonmotile, nonsporeforming soil diptheroids,

Fig. 6. Sample Sites 122 and 123 (patches of moss and lichen-encrusted soil between macro vegetation)

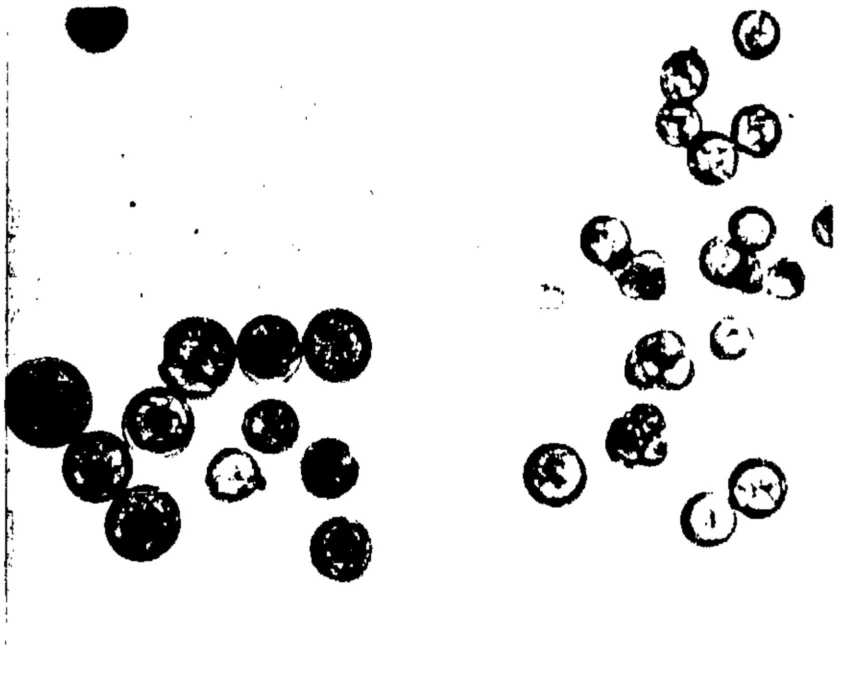


Fig. 7. *Protococcus grevillei*, X 500 (a coccoid green alga from Soil 115)

Fig. 8. *Penicillium* sp., X 1,250 (conidia, sterigma, and condidiophore, lactophenol stain)

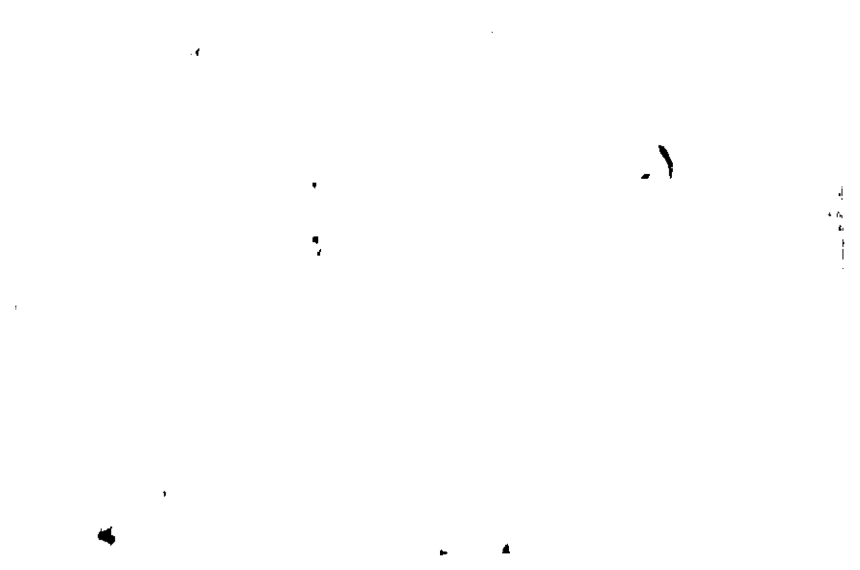


Fig. 9. *Bacillus macerans*, X 1,250 (gram variable spore-forming peritrichous rods from Soil 118)

Corynebacterium sp., and (2) sporeforming, peritrichous *Bacillus* sp., e.g., *B. macerans* (Fig. 9), *Achromobacter* sp., *Pseudomonas* sp., and *Brevibacterium* sp. (Ref. 7). Nearly all are microaerophilic bacilli. A common sulfur-iron organism which was isolated from a microaerophilic medium is *Thiobacillus coproliticus* (Fig. 10).

4. Summary and Conclusions

In many ways the Valley of 10,000 Smokes presents a unique, harsh environment which limits the numbers and kinds of microorganisms present. Some of the organisms present are typical of arid desert areas previously investigated (Ref. 7 and Footnote 1, p.198), but others are more specialized and restricted to unique volcanic habitats.

Fig. 10. *Thiobacillus coproliticus*, X 1,250 (microaerophilic monotrichous, filamentous, autotrophic sulfur oxidizer from Soil 119)

Limited numbers of organisms can only be accounted for by a number of soil, climatic, and other local environmental factors. The soils themselves were not toxic, as evidenced by lack of inhibition when wet Valley soils were applied to plate cultures of *Bacillus subtilus var. niger* and *Escherichia coli*. Restrictive soil physical factors influencing microorganisms in the Valley included a rapid rate of surface desiccation, fast infiltration and percolation rate, rapid leaching, and a narrow range of available soil moisture. Restrictive soil chemical factors included an extremely low cation exchange capacity, a very low or negligible nutrient availability status, and absence of clay minerals. Low pH, little if any inorganic sources of carbon or nitrogen, and very little organic matter also contributed to a sparse microflora. Proximity of vegetation definitely influenced total numbers and kinds of organisms present. In general, the further the distance from vegetation the fewer the total number of microorganisms and the fewer algae. The most unfavorable soils for microorganisms, which were in the center of the Valley, showed a pronounced glassy phase and mineralogically resembled perlite. Primary micrometeorological factors unfavorable to growth of microorganisms included high velocity, sulfurous, ash, and pumice-laden winds which abraded and desiccated the Valley surface. There has been no evidence of gases issuing from an underlying body of magma in the Valley (Ref. 8). Emission of volcanic gases upward through the soil would undoubtedly influence not only the distribution and numbers of microflora, but also determine the kinds of indigenous soil microorganisms present.

B. Soil Studies—Microflora of Desert Regions. IV. Soil Extract as a Culture Medium

R. E. Cameron, F. A. Morelli, and G. B. Blank

1. Introduction

Nearly all of the life detection experiments for Mars will sample dust, soil, or soil-like materials (Ref. 9). Two of these experiments, "Gulliver" (Ref. 10) and "Wolf-Trap" (Ref. 11), will contain previously prepared media for culturing microorganisms which may occur in the Martian sample secured from the planetary surface.

There is no synthetic "universal medium" for soil organisms. It is known that specific populations of soil microorganisms can have highly selective nutritional requirements. Therefore, it is possible that Martian microorganisms may not be able to utilize any of the prepared media included in extraterrestrial life detection experiments. Both the "Gulliver" and "Wolf-Trap" are limited as to kinds and amounts of culture media, and only one or several media can presently be included in the experiments.

The most optimum culture media for unknown Martian soil microorganisms consequently present a critical parameter for media-dependent life detection experiments.

2. Soil Extract Media

One solution to the problem of utilizing the proper medium for culturing unknown extraterrestrial microorganisms involves the acquisition and preparation of an extract from the Martian soil. This will insure a greater probability of including available or soluble inorganic and organic components required by Martian microorganisms. A modification of this Martian soil extract medium could include the addition of previously prepared simple sugars, such as glucose, and broad spectrum proteins composed of microbial cells, such as yeast extract. Both glucose and yeast extract promote or stimulate the metabolism, growth, and reproduction of many terrestrial soil microflora, especially when both are used in culture media. Specific inorganics, e.g., sulfur, or sodium thiosulfate, stimulate soil microorganisms that are especially fastidious, e.g., sulfur metabolizers, and could also be included with soil extract media.

Having previously observed the effectiveness of *properly prepared* soil extracts (Ref. 6), it is the purpose of this article to present a preliminary experiment on the comparative effectiveness of soil extract, and soil extract enriched with yeast extract, as growth media for soil microorganisms. Soil and yeast extract were prepared according to detailed procedures given previously (Ref. 6).

3. Experiments

a. Soil. An ancient beach soil near the San Andreas Fault was collected from the top 5/16 in. of the surface in the northern Salton Sea area of the Colorado Desert in October, 1962 (Fig. 11). A report on the detailed description of the soil site, and the physical, chemical, and microbiological characteristics of this soil and methods of analysis is in preparation.[3]

Physically, this soil is a thin, light-gray, structureless algal-lichen surface crust of medium porosity and bulk density, and low water activity (Fig. 12). Mineralogically,

[3] Cameron, R. E., Morelli, F. A., and Blank, G. B., *Desert Soil Characteristics. I. Preliminary Studies*, Technical Report, Jet Propulsion Laboratory (to be published).

Fig. 11. Collection of soil crusts in Colorado Desert, California

Fig. 12. Thin, algal-lichen surface crusts used for soil extract media

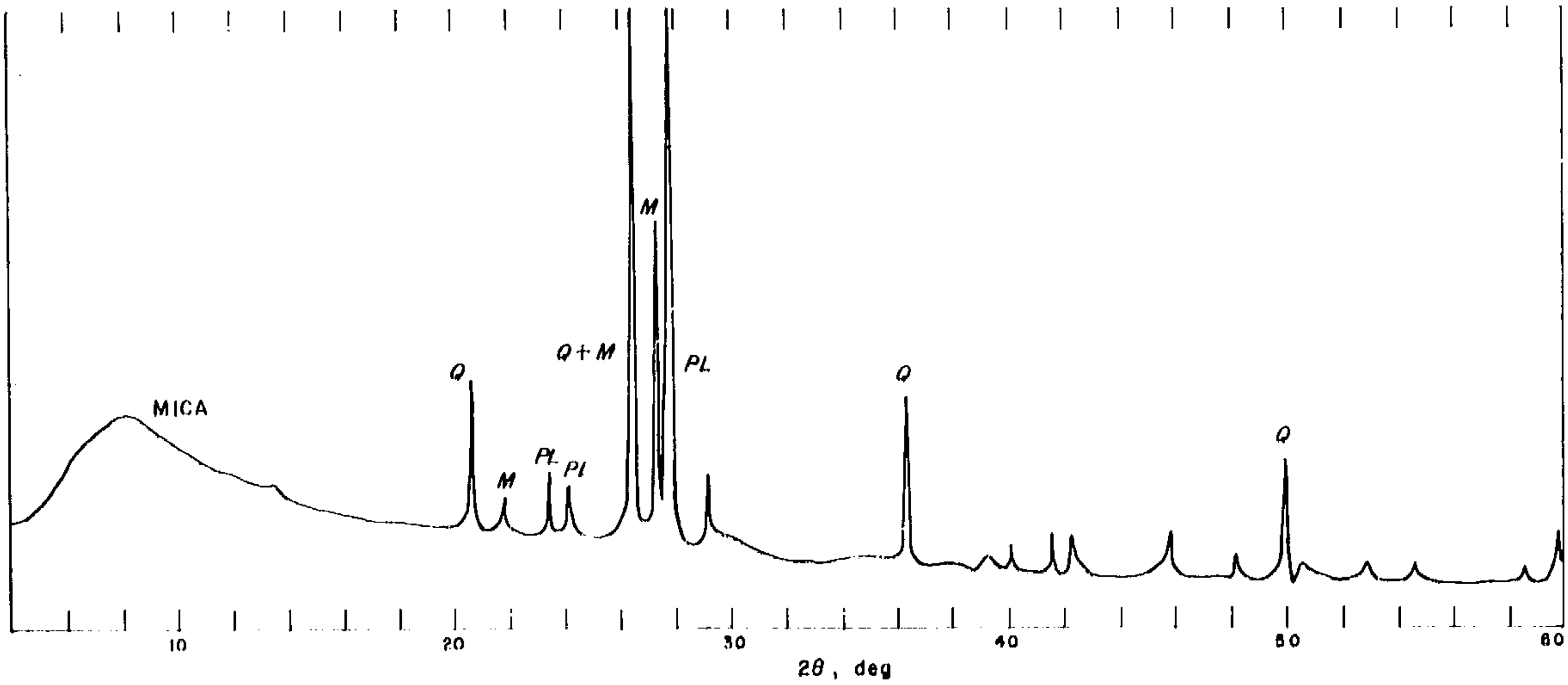


Fig. 13. Smoothed X-ray diffraction curve for soil used to prepare soil extract media (Q = quartz, *Pl* = plagioclase, *M* = microcline)

it is composed of quartz, plagioclase, microcline, and mica (Fig. 13). Chemically, on a weight basis, this soil contained 0.15% protein, 674 ppm fats, and 0.16% carbohydrates. Total carbon and total nitrogen were 0.44% and 0.042%, respectively, with a ratio of 10.5, which is typical for decomposed soil organic matter and microbial cells. Saturated soil paste values showed a pH of 7.9, an Eh of +375 mv, an electrical conductivity of 29.6 × 10^5 mhos/cm at 25°C, and an osmotic pressure of 0.009 atm. A low cation exchange capacity, 6.0 meq/100 g of soil, was primarily occupied by Ca^{++} as $CaCO_3$. Microbiological determinations showed the presence of 3.7 × 10^5 aerobic bacteria per gram of soil. Three hundred molds and at least 10^3 algae per gram of soil were also present.

b. Growth measurements. Three series of flasks with 4 flasks in each series were designed for culture and nephelometry[1] (Fig. 14). Each flask contained 100 cm³ of culture medium. The first series contained desert soil extract (0.15% extractable organic matter), the second contained soil extract plus 0.05% yeast extract, and the third contained only 0.05% yeast extract. Half of the flasks were inoculated with 22 viable *Bacillus subtilus* spores/cm³ into 100 ml of media. Mixed populations of indigenous soil microflora containing 370 aerobic bacteria (plus actinomycetes) and 3 fungi/cm³ were added to the remaining flasks containing 100 cm³ of media. All cultures were incubated at 25°C in the dark without shaking until the termination of the experiment after 92 to 118 hr.

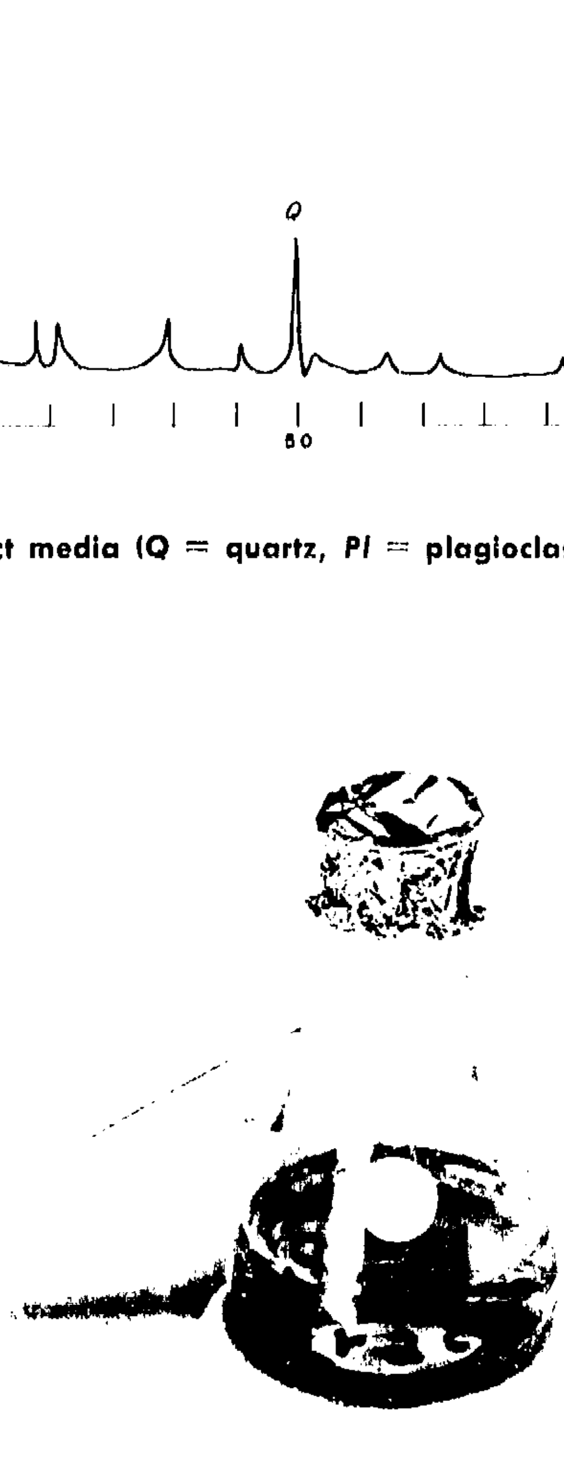


Fig. 14. Nepheloflask used for culture and determination of optical density of microorganisms

[1] Heyn, H., Engineering Assistant, Jet Propulsion Laboratory, Chemistry Section.

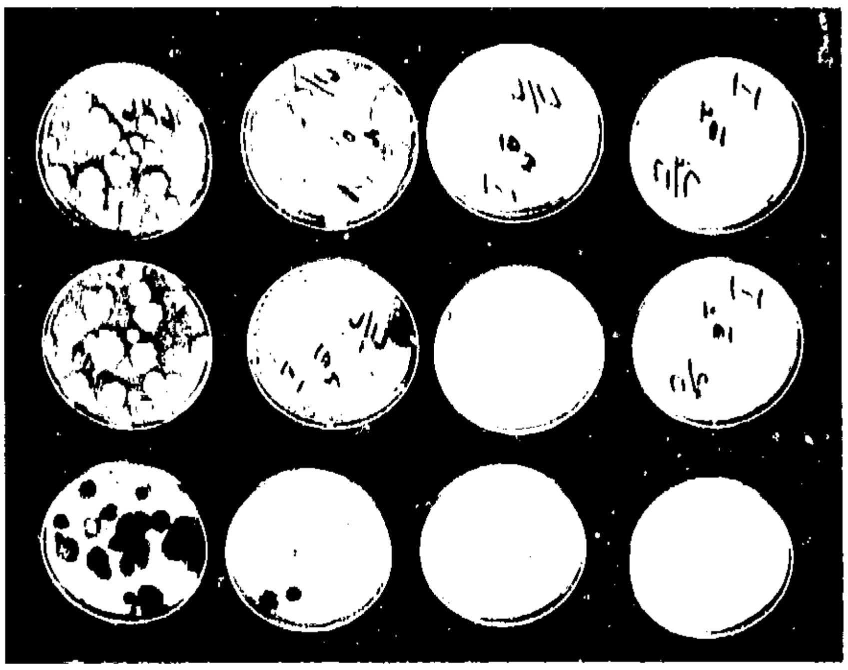


Fig. 15. Dilution plates used for determining number of bacteria and fungi in soil and cultures (upper two rows of plates are white bacteria colonies on trypticase soy agar; lower row of dark colonies shows molds on rose bengal agar

Control cultures contained the same media plus killed organisms.

All measurements were performed in duplicate. Growth was determined with time by means of (1) optical density using a Klett-Summerson photoelectric-colorimeter; (2) plate counts on trypticase soy agar for aerobes, and rose bengal agar for fungi (Fig. 15); (3) electrical conductivity; (4) pH; and (5) Eh. A comparison of (1) soil, (2) soil minus extract, (3) soil extract, and (4) yeast extract was obtained by means of mass spectrometry, differential thermal analysis, and infrared absorption spectrometry[5], and will be reported later.

4. Results and Discussion

For the purpose of this article, results are presented only for measurement of growth with time by means of aerobic plate counts and optical density (Figs. 16, 17, and 18). There was a definite lag phase before growth was detected by either plate counts or optical density measurements. This was partly due to the extremely small numbers of microorganisms inoculated into the flasks. However, following the initial lag phase of $>$ 17 but $<$ 40 hr, there was a steep rise in growth, except for *Bacillus subtilus* which did not grow in the medium containing yeast extract. Growth curves generally showed a remarkable similarity (especially Figs. 16 and 18). Short lag or depressive phases in growth were also noted during the progress of the experiment. However, this may have been an "apparent" lag or depressive phase of fewer bacteria because there were clumps of organisms, inhibition due to toxic-formed substances, settling of organisms, and adherence to the inner surface of the culture flask. Bacteria were not grown in the attached nephelotube, and therefore did not give a false rise in optical density readings because of tube adherence. Dilution plates and microscopic examination indicated that many of the bacteria were flagellated bacilli that rapidly formed spreading confluent colonies on trypticase soy agar. As indicated by plate counts as well as optical density measurements, maximum growth was generally not obtained by the end of the experiment.

Following the end of the experiment, the cultures were exposed to the light. They developed prolific growths of filamentous blue-green and coccoid green soil algae in soil extract and soil extract plus yeast extract. Soil algae did not develop in yeast extract. Growth of fungi was also noticeable in yeast extract.

5. Conclusions

On the basis of preliminary results, it has been shown that a desert soil extract, or the same desert soil extract enriched with yeast extract, is an excellent medium for growth following extremely low inoculations of *Bacillus subtilus* as well as a very low inoculum of mixed populations of microorganisms indigenous to the soil from which the extract had been prepared. Variable results are obtained with yeast extract medium alone, and it may not be a suitable medium for a single population of microorganisms, such as *Bacillus subtilus*, which is an abundant inhabitant of many typical agricultural soils. Additional desert soil extract should be prepared and tested with microflora from the same soils and different soils used to prepare the extract. More complete testing of soil extract enrichment with various organics and inorganics should be undertaken. Single and mixed populations of desert soil microorganisms should also be tested for growth responses under different environmental conditions, e.g., temperature, osmotic and barometric pressure, gases, and light, in various combinations of soil extract and soil extract enrichment media.

[5] Cameron, R. E., Kuwada, D. M., Blank, G. B., and Morelli, F. A., "Soil Organic Matter Characterization," Jet Propulsion Laboratory (unpublished results, 1963).

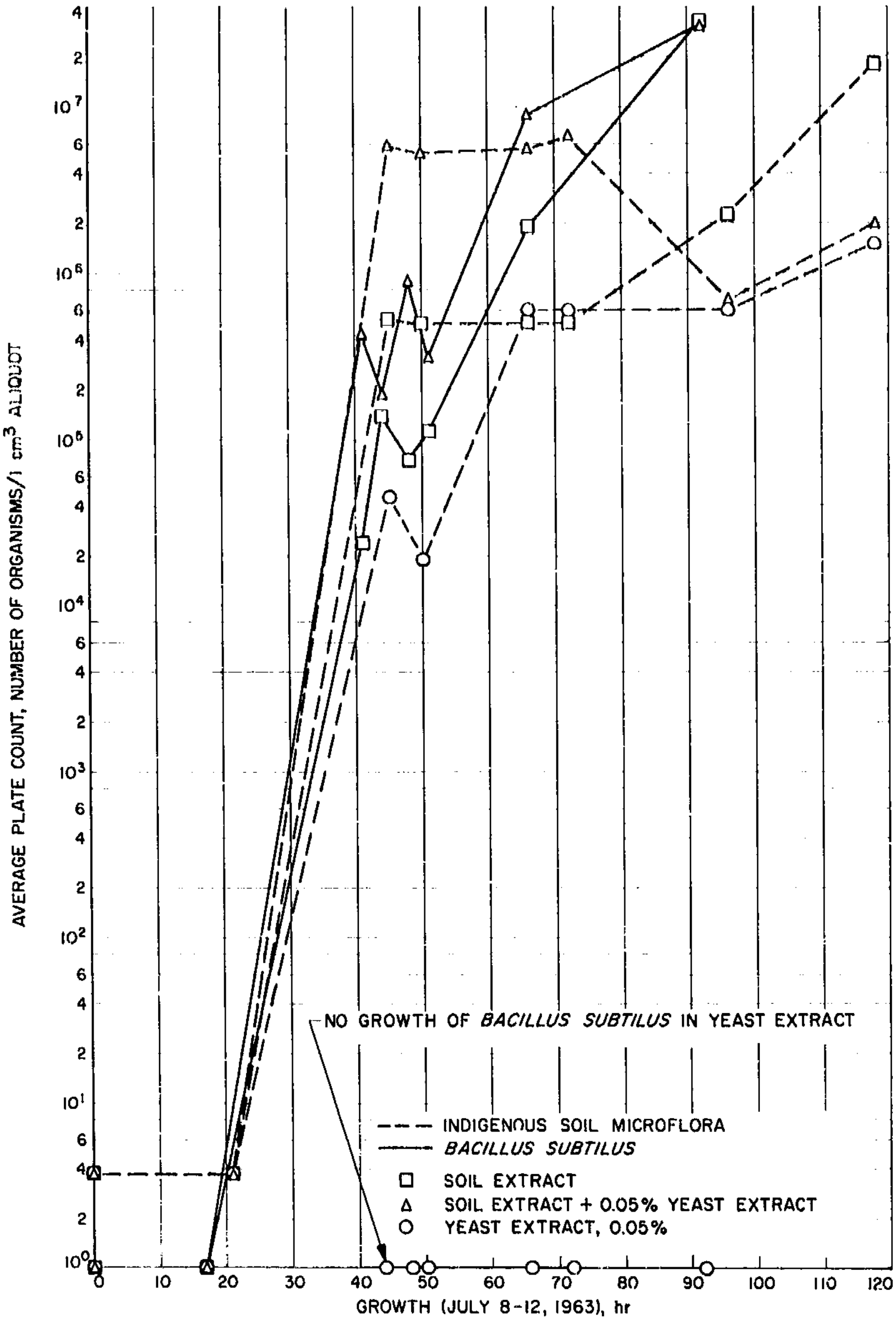


Fig. 16. Growth curves of aerobic bacteria obtained by dilution plate counts

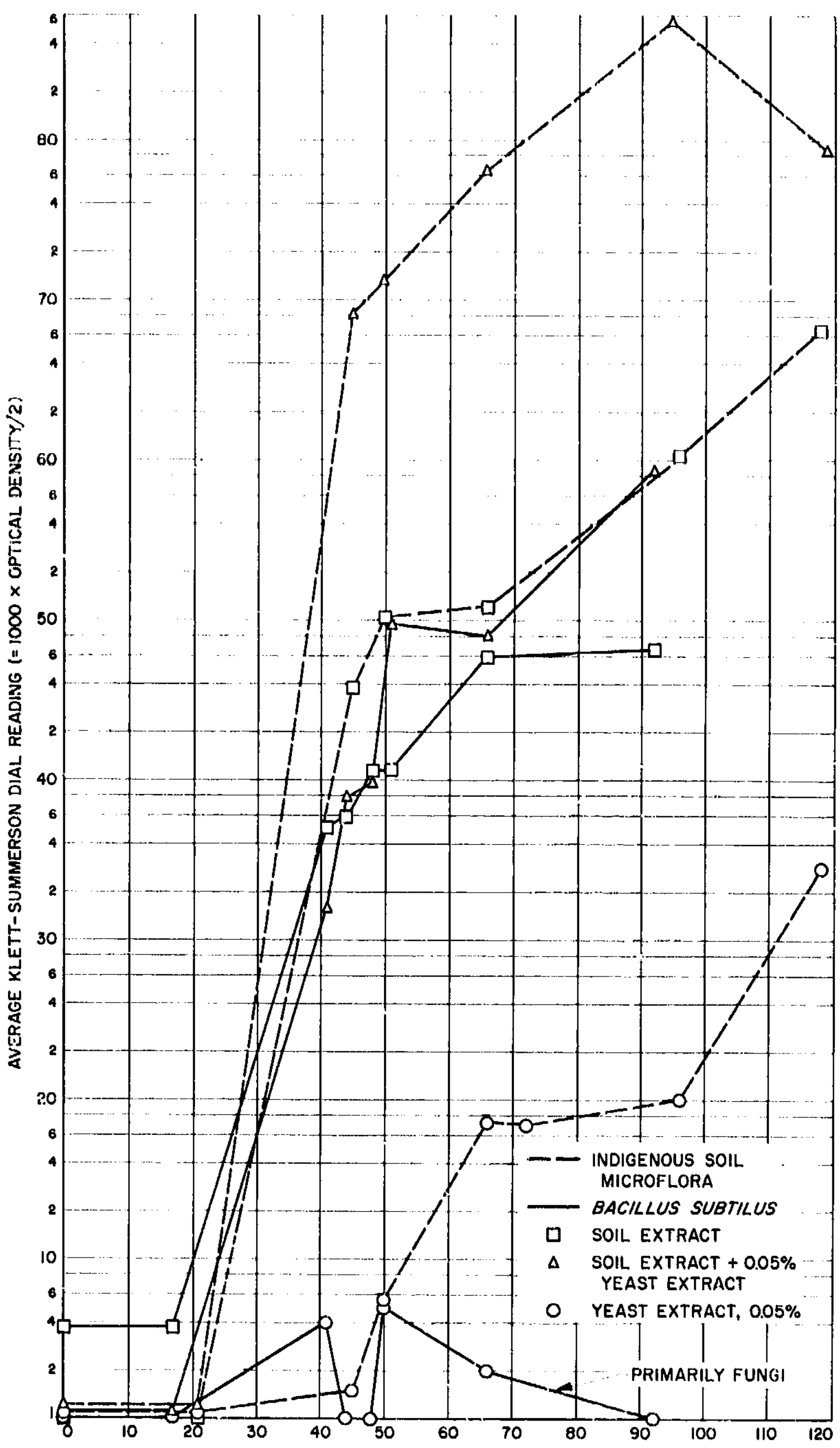


Fig. 17. Growth curves of aerobic bacteria obtained by optical density measurements

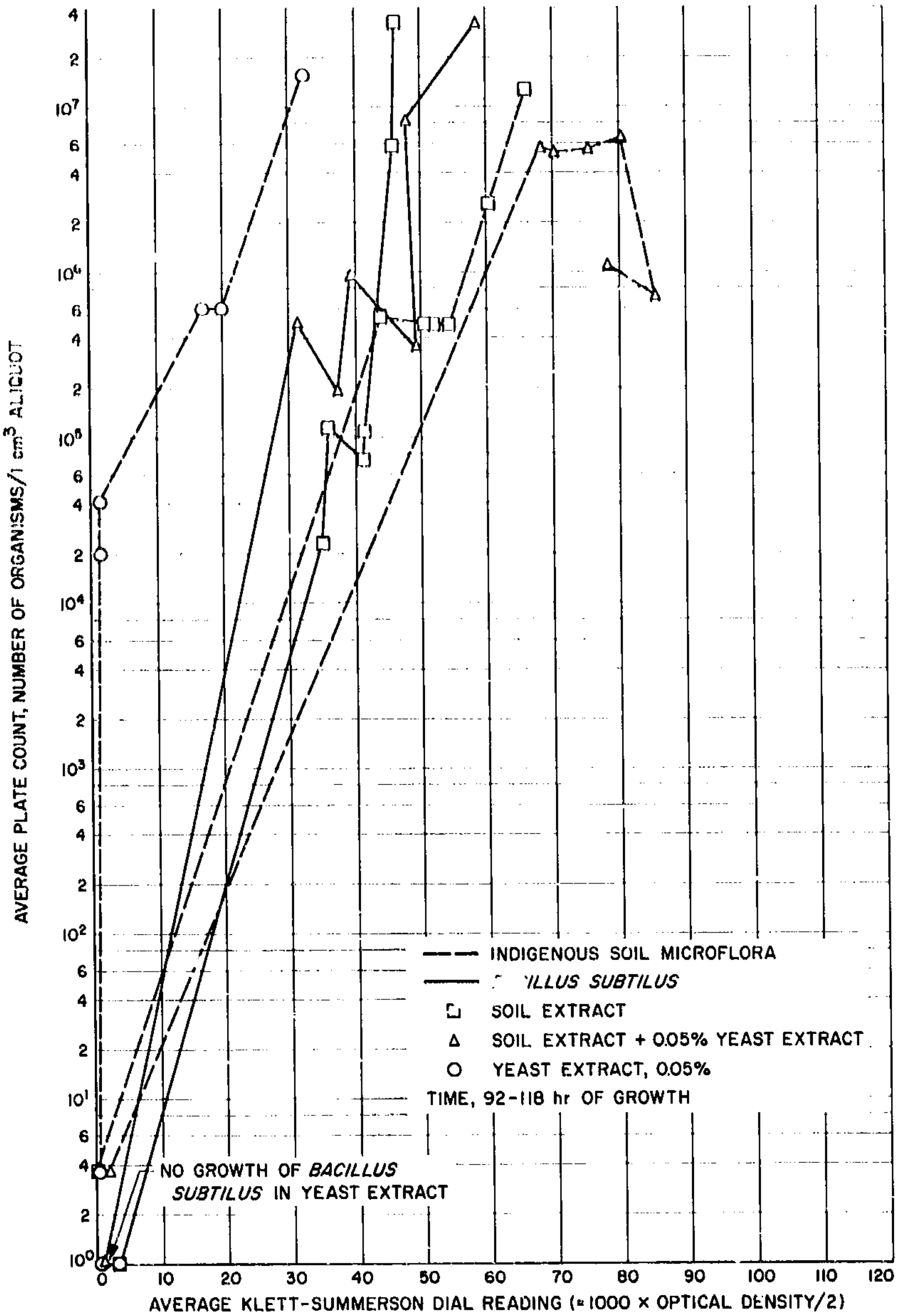


Fig. 18. Correlation of growth by dilution plate counts and optical density

C. Soil Studies—Desert Microflora. V. Soil CO_2 Production Measured by Gas Chromatography

G. B. Blank, F. A. Morelli, R. E. Cameron, and G. L. Hobby

1. Introduction

The production of CO_2 in soils can serve as a useful means of indicating biological activity, and indirectly give an estimate of the quantity or volume of microflora present in a given sample of soil. This experiment was conducted in order to determine the production of CO_2 in a desert soil of very low carbonate carbon content after activation of the microflora by the addition of water. When a soil is air dried, and then remoistened, there is rapid CO_2 evolution (Ref. 12). Major factors influencing the evolution of CO_2 are the organic matter levels of the soil, the kinds of organic matter and its state of decomposition, the C:N, depth of soil, aeration, gas concentrations, temperature, water activity, pH, and Eh, as well as the numbers and kinds of microorganisms.

2. Description of Portable Gas Chromatograph (GC)

The GC apparatus[6] used in this preliminary experiment contained a thermal conductivity detector and four gas separation columns. One "preseparation" column could be arranged in series with any one of three individual columns. No column heaters were included in this unit, and gas separations were achieved at room temperature[6]. The column used for the preliminary CO_2 experiment reported here was 15 in. in length, and was packed with "unactivated" silica gel, 160/170 mesh. The entire experiment was conducted at a room temperature of 20–22°C.

3. Materials and Methods

a. Soil description. Soil 76-1, used in this preliminary experiment, was obtained from the Mohave Desert, northwest of Mohave, California, at an elevation of 2,760 ft. It was collected in October 1963, and stored in the air-dry state until used for this experiment. The sample was taken from a depth of ¼ to 6 in. beneath the soil surface (Fig. 19). Texturally, this soil was classified as a sand, having 84.1% sand, 9.6% silt, and 6.3% clay. A bulk density value of 1.57 g/cm^3 and a porosity value of 41.1% were determined, and these values are typical for a sandy soil. At field capacity (0.5 bar moisture stress), the pH of the soil was 7.5 and the Eh was +458 mv. The cation exchange capacity was 7.0 meq/100 g of soil. This soil contained 0.9% organic matter, 0.55% organic carbon, 0.03% carbonate carbon, and 0.060% nitrogen, nearly all of which was in the organic form. A C:N ratio of 11.1 was indicative of decomposed organic matter and/or microbial cells. A complete report is in preparation for details of soil properties and methods of soil analysis.[7]

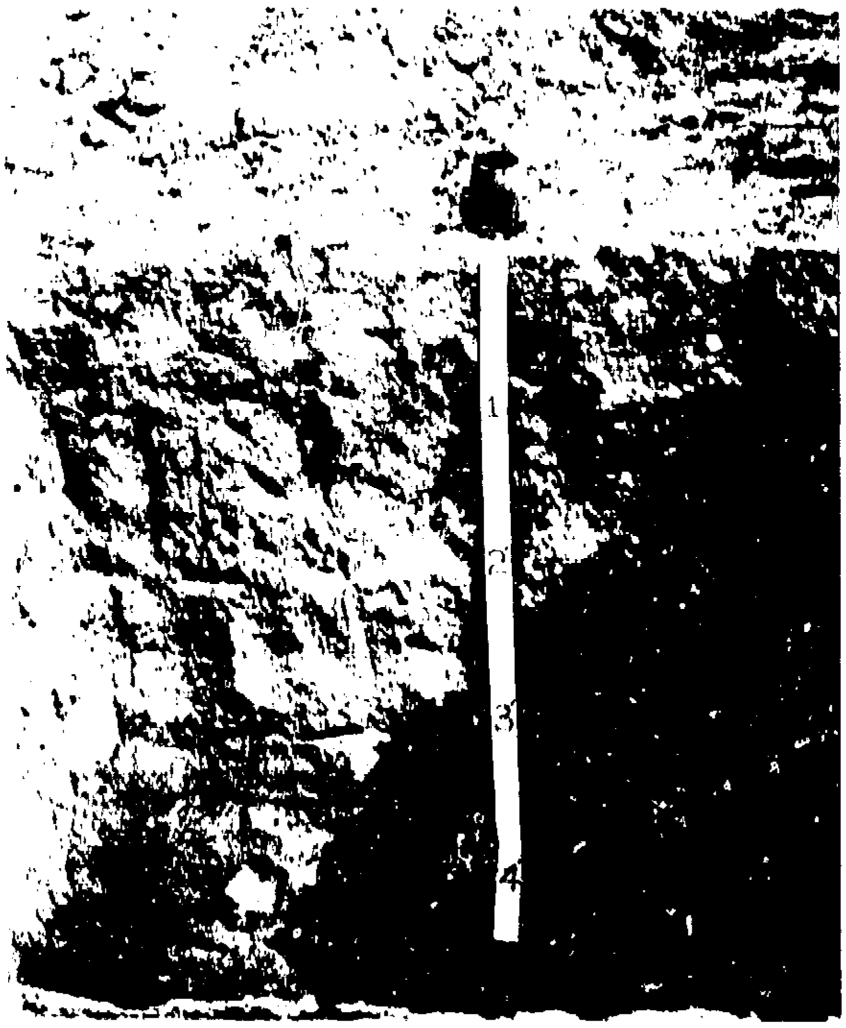


Fig. 19. Soil profile pit containing Soil 76-1 at ¼- to 6-in. level

b. Sample preparation and analyses. Samples of air-dry soil with particles <2 mm were packed into 6 open glass tubes 1 in. long × ¼ in. in diameter with a volume of 0.66 cm^3. Soil sample weights were 0.90 g. Rubber septums were placed on each end of the tubes.

Gas-tight syringes with very fine hypodermic needles were used to extract 25 μl of soil air. Sixty μl of water were added to each tube to bring the soil to field capacity

[6]Loenco, Model 15B, Gas Chromatograph Operation Manual; Stembridge, C., "Portable Gas Chromatograph," *Space Programs Summary*, Jet Propulsion Laboratory (to be published).

[7]Cameron, R. E., Morelli, F. A., and Blank, G. B., *Desert Soil Characteristics. I. Preliminary Studies*, Technical Report, Jet Propulsion Laboratory (to be published).

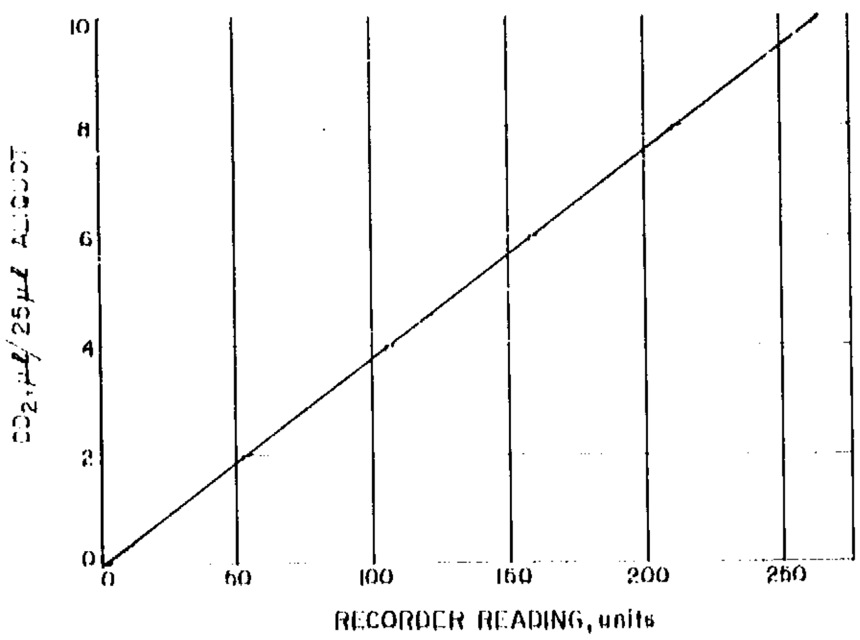


Fig. 20. CO_2 calibration curve on "unactivated" silica gel, 160/170 mesh, obtained at room temperature (20 to 22°C) with detector current of 115 ma

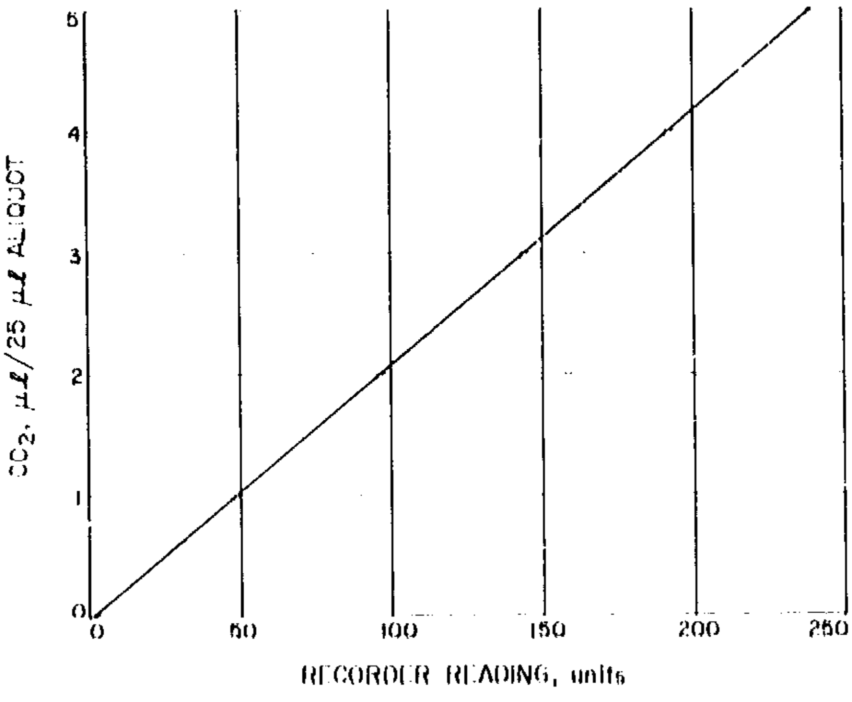


Fig. 21. CO_2 calibration curve on "unactivated" silica gel, 160/170 mesh, obtained at room temperature (20 to 22°C) with detector current of 130 ma

(0.5 bar). Each sample was incubated for more than 400 hr at room temperature. Phenol and $HgCl_2$ were injected into duplicate samples after 200 hr so as to observe their effect upon the inhibition of microorganisms and any subsequent changes in release of CO_2. Control samples consisted of dry soils in similar sample tubes. Hydrochloric acid was added to determine its effect upon any carbonates present in the soil (dry samples). At the termination of the experiment, dilution plate counts were made on soil samples for (1) aerobic bacteria, plus actinomycetes, and (2) anaerobic bacteria.

For this experiment, the He carrier gas was passed through the columns at a pressure of 30 psi, and with a flow rate of 55 ml/min. In the first part of this experiment, the CO_2 was analyzed with a detector current of 115 ma, but increased sensitivity was desirable for measuring small changes in CO_2 concentration. Sensitivity was then doubled for the latter part of the experiment by raising the detector current to 130 ma (Figs. 20 and 21).

4. *Results*

Changes in concentration of 0.05 μl of CO_2 could be detected. When 25 μl of soil air were taken from the wet soils (Fig. 22), a noticeable increase in CO_2 production occurred in all samples after 3 hr. After 180 hr, the CO_2 concentration had increased from 0.03 to 0.25 μl/25 μl of soil air sample. The increase in CO_2 from the addition of 60 μl of 0.1N HCl to air dry controls never exceeded more than 0.075 μl CO_2.

The addition of 60 μl of 1% $HgCl_2$ after 120 hr resulted in a rapid evolution of CO_2. Concentrations of detectable CO_2 increased from 0.2 μl/25 μl aliquot to almost 0.6 μl/25 μl in less than a few hours (Fig. 22).

One hundred and fifty hr after the addition of phenol, the CO_2 concentration in these soil tubes increased to 0.1 μl/25 μl, and then decreased to control levels at the end of the 370-hr incubation period.

Although both phenol and $HgCl_2$ are biological inhibitors, phenol decreased CO_2 formation whereas $HgCl_2$ led to a rapid CO_2 evolution. Mercuric chloride apparently released CO_2 sorbed or mineralized in the soil matrix by acidification of the samples. Phenol may have killed soil microorganisms, but it did not result in a rapid release of CO_2.

Average dilution plate counts of duplicate samples of aerobes and anaerobes are shown in Table 7. These microorganisms were cultured on trypticase soy agar, incubated at 22°C, and the anaerobic plates were placed in a CO_2 atmosphere. Values obtained on dry control soils were less than those obtained on soils wet with water for more than 400 hr. The lowest counts were obtained on dry soils which had HCl added to them, and low counts were also obtained on wet samples containing added $HgCl_2$. Although counts could not be obtained during the progress of the experiment, it is evident that there was a considerable number of survivors remaining at the conclusion of this experiment.

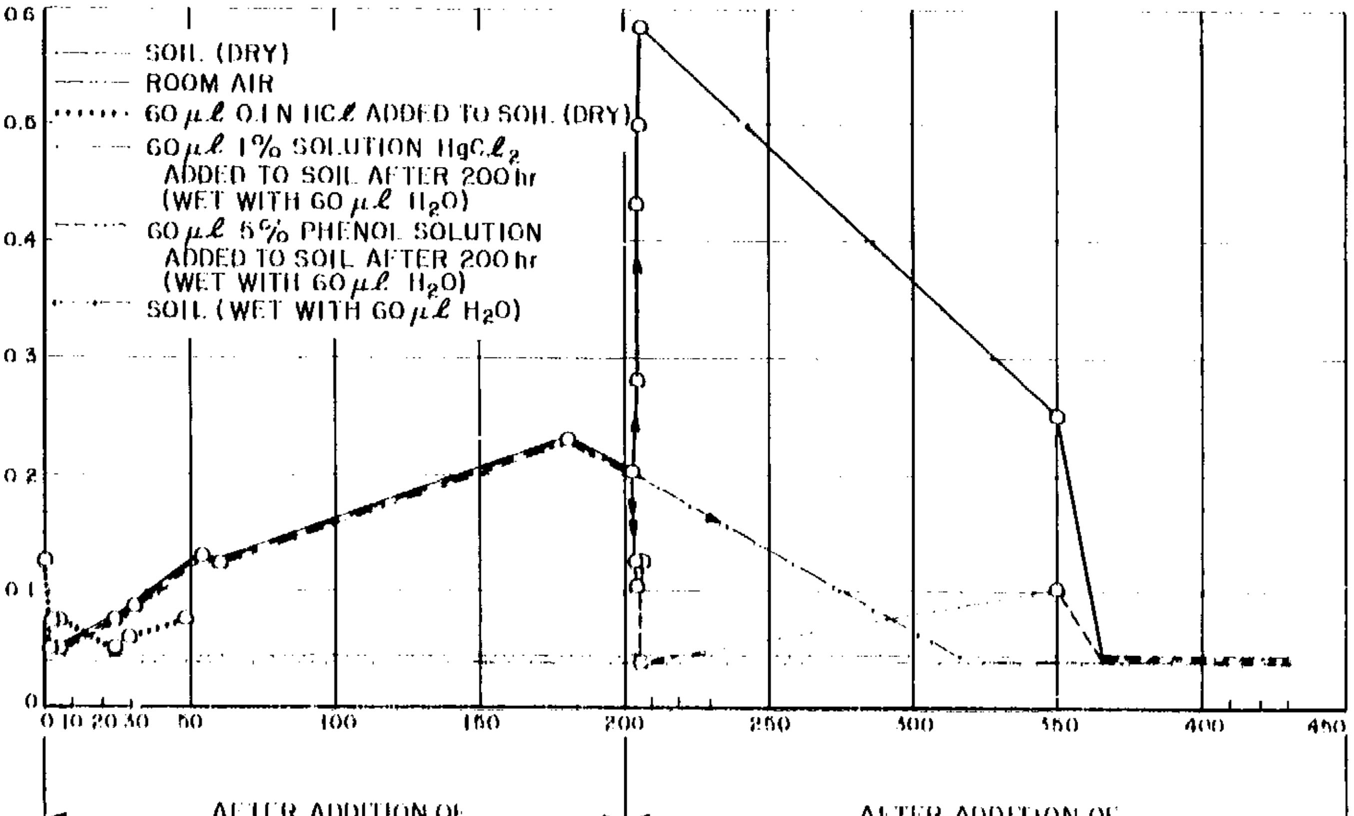


Fig. 22. CO_2 evolution by desert soils

5. Conclusions and Future Experiments

A gas chromatograph operated at room temperatures (20–22°C) with the proper type of gas column characteristics is capable of separating and measuring CO_2 production in the soil. The soil used in this experiment contained very little natural carbonate, and CO_2 production was probably caused by the activities of the indigenous soil microflora.

In future experiments, research will be undertaken with sterilized and ignited soils, soils having different physical, chemical, and microbiological properties, soils containing added carbonates, and soils inoculated with soil extract, glucose, and yeast extract, in order to determine the effectiveness of this method for stimulating CO_2 evolution by desert soil microorganisms. Variations in O_2 concentration will also be measured with time, as

Table 7. Numbers of aerobic bacteria plus actinomycetes and anerobic bacteria in soil samples

Sample number	Treatment	Microorganisms/g of soil (average): Microaerophils (facultative anaerobes)	Anaerobes	Aerobes (bacteria and actino-mycetes)
1	Air-dry, control	10^7	3.3×10^4	3.2×10^6
2	Air-dry, control	10^7	6.2×10^3	3.9×10^6
3	Air-dry, 0.1N HCl	10^7	1.4×10^4	7.7×10^5
4	Air-dry, 0.1N HCl	10^7	4.0×10^4	5.5×10^5
5	Wet, control	10^7	4.2×10^4	15.2×10^6
6	Wet, control	10^7	3.3×10^4	9.7×10^6
7	Wet, 1% $HgCl_2$	10^7	6.6×10^3	2.5×10^5
8	Wet, 1% $HgCl_2$	10^7	3.3×10^3	6.0×10^5
9	Wet, 5% phenol	10^7	6.6×10^3	2.5×10^5
10	Wet, 5% phenol	10^7	No growth obtained	2.7×10^6

well as indications of production of CH_4, CO, and H_2, which occur when organic carbon is incompletely metabolized under anaerobic conditions. A column has already been selected and calibrated for this purpose.[8]

D. Soil Studies—Desert Microflora. VI. Abundance of Microflora in an Area of Soil at White Mountain Range, California

R. E. Cameron, F. A. Morelli, and G. B. Blank

1. Introduction

Sampling a planetary surface for extraterrestrial life is a critical factor for life detection purposes. Among the factors that must be considered are selection of the sample site, the method of sample procurement, and sample processing or sizing prior to tests for life detection. Samples proposed for use with extraterrestrial life detectors include devices depending on adhesives or sticky strings, aerosol or vacuum-type collection, and electrostatic precipitation (Refs. 1 and 13). Some samplers obtain materials from a wide area, whereas others are more restricted to a limited area. All of the currently designed samplers will be subject at least to some extent on ambient environmental factors at the Martian surface.

Martian "soil" may or may not be homogeneous over a wide area and volume of planetary material. Probably, because of aridity, it will not contain as many favorable microenvironments for organisms as are found in terrestrial soils. On our own planet, soils differ in properties from point to point within a given volume and area of material, especially in typical fertile, productive agricultural soils which have developed optimally under favorable environmental influences. In arid and semi-arid soils, variabilities in soil properties within a given ecosystem are not known. The purpose of this article therefore is to determine the variability of microflora in a given area of surface soil of the same genesis in a comparatively arid area.

[8] Stembridge, C., "Portable Gas Chromatograph," *Space Programs Summary*, Jet Propulsion Laboratory (to be published).

2. Materials and Methods

Eighteen samples were randomly collected from the surface 1 in. of soil within a protected, alpine field plot[9] (Figs. 23 and 24). All of the soils were loamy sand in texture. Saturated pH values were 6.1 to 7.0. The mineralogical composition of these soils was silica, plagioclase, feldspar, microcline, and mica.

[9] White Mountain Research Station, Dr. Nello Pace, Director, University of California, Berkeley, California.

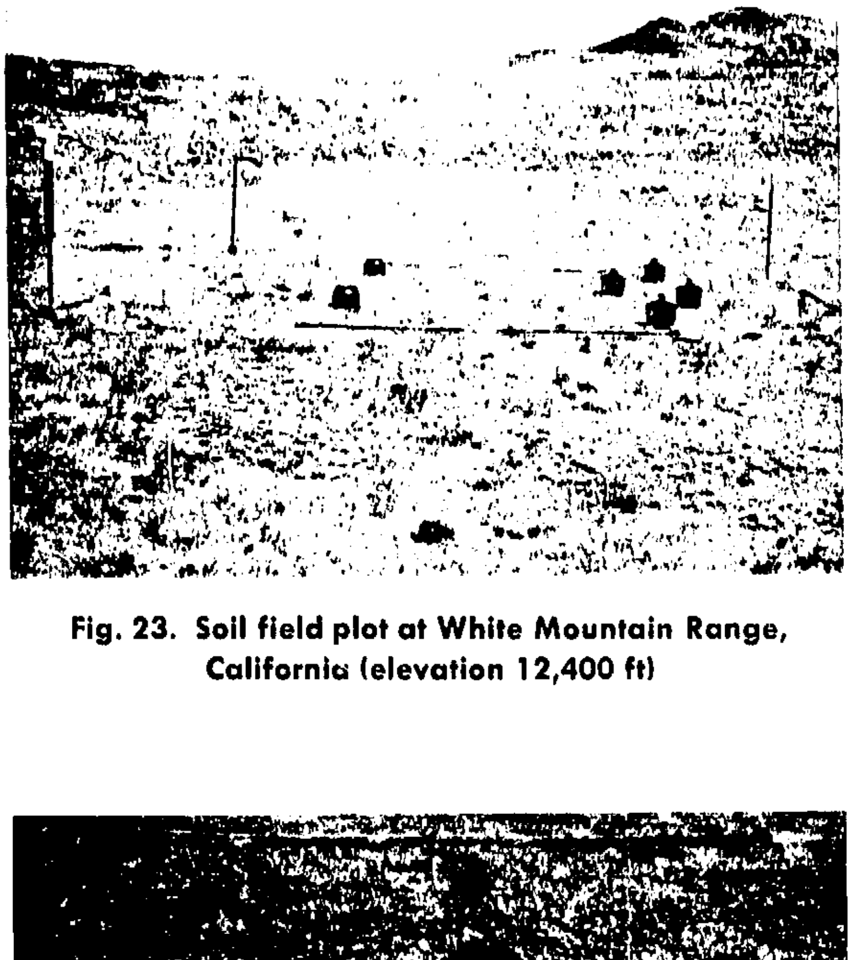


Fig. 23. Soil field plot at White Mountain Range, California (elevation 12,400 ft)

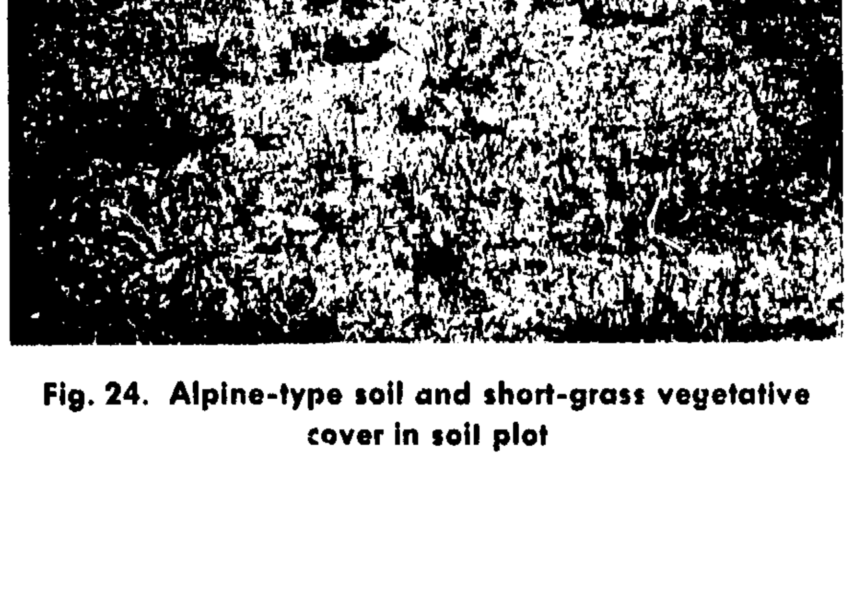


Fig. 24. Alpine-type soil and short-grass vegetative cover in soil plot

Ten-gram quantities of air-dried samples were suspended in 90 grams of water. This suspension was diluted 1/10. Appropriate aliquots of soil-water suspension were then pipetted into selective media. For the various microflora this included the following media: (1) aerobic bacteria (including actinomycetes), trypticase soy agar, (2) microaerophilic bacteria, fluid thyoglycollate medium, (3) anaerobes, trypticase soy agar in CO_2, (4) fungi (molds), rose bengal agar, and (5) algae, 8 parts Pochon inorganic medium plus 1 part 1:1 desert soil extract. All bacteria and fungi cultures were incubated at room temperature of 22°C for at least 2 weeks; algae were cultured in warm white fluorescent light (450 to 550 ft-c) for at least 4 months. Details of media and culture conditions will be presented in a forthcoming report[10].

[10]Cameron, R. E., Morelli, F. A., and Blank, G. B., *Desert Soil Characteristics. I. Preliminary Studies*, Technical Report, Jet Propulsion Laboratory (to be published).

3. Results and Discussion

After the termination of incubation periods, average plate counts were made for the aerobic and anaerobic bacteria and fungi (Table 8). Presence of growth or no growth of the microaerophilic bacteria and algae are also shown for dilution cultures (Table 8). A statistical analysis is given for each group of microflora as indicated by the average, standard deviation, standard error, and range of organisms.

The aerobic and microaerophilic bacteria were the most abundant group of microflora present in all of the 18 samples. The next most numerous groups present in these samples were the algae and the anaerobic bacteria. The fungi were the least abundant group of microorganisms present. Values obtained for microflora in these 18 samples were similar to those obtained for numbers of

Table 8. Abundance of microflora at White Mountain soil plot

Soil 9-2 sample number	Aerobes (bacteria + actinomycetes) $\times 10^3$/g soil	Microaerophiles growth in dilutions/g soil[a]		Anaerobes $\times 10^3$/g soil	Fungi (molds) $\times 10^3$/g soil	Algae growth in dilutions/g soil[a]			
		10^6	10^7			10^3	10^4	10^5	10^6
1	1150	+	—	670	3.4	+	+	+	
2	1450	+	—	500	3.5	+	+	+	+
3	6900	+	—	370	4.0	+	+	—	
4	1600	+	—	35	5.2	+	+	—	
5	1080	+	—	68	4.9	+		—	
6	730	+	+	600	2.9	+	—	—	—
7	1650	+	+	127	2.2	+	+	—	—
8	7500	+	+	37	2.6	+	+	—	
9	1040	+	—	150	1.7	+	—	—	—
10	2200	+	+	58	2.7	+	+	—	—
11	4400	+	+	660	12.8	+	+	+	—
12	830	+	+	230	2.7	+	+	+	—
13	1290	+	+	8.3	39.0	+	+	+	—
14	760	+	+	7.5	12.4	+	+	+	—
15	610	+	+	38	8.8	+	+	—	—
16	1160	+	+	0.13	4.3	+	+	—	—
17	730	+	+	81	3.3	+	+		—
18	670	+	+	49	2.6	+	+	+	+
Average	1986	7×10^6		204.94	6.61	144×10^3			
Standard deviation	2089	4.4×10^6		241.08	8.72	314×10^3			
Standard error	1451	4×10^6		200.04	5.17	1903×10^3			
Range	$610 \times 10^3 \rightarrow 7500 \times 10^3$	$10^6 \rightarrow 10^7$		$0.13 \times 10^3 \rightarrow 670 \times 10^3$	$1.7 \times 10^3 \rightarrow 39 \times 10^3$	$10^3 \rightarrow 10^6$			

[a]—, no growth
+, growth at highest dilution

microflora in Southwestern United States desert soils[11]. In general, abundancy ratios for six California surface desert soils were as follows: microaerophilic bacteria > aerobic bacteria > algae > anaerobic bacteria > fungi. Not all of these groups are routinely determined in agricultural soils, nor have the culture media utilized for this study been extensively used for a study of soil microflora.

In well-developed agricultural soils, the most abundant group of microflora present is the aerobic bacteria. Anaerobic bacteria are the next most numerous, and fungi are commonly the third most abundant group of microflora (Ref. 14). However, the fungi may predominate in an agricultural soil, depending on local (microenvironmental) conditions. Numbers of algae appear to be lower in typical well-developed soils than in arid or semi-arid soils (Ref. 14). Numbers of microaerophilic bacteria are not commonly reported for soils.

There was little deviation in the numbers of aerobes and microaerophiles obtained from the 18 samples in our soil field plot. More variability was shown for numbers of anaerobes and algae. One soil, Sample 16, had only 0.13×10^3 anaerobes/g of soil, whereas the average was 204.9×10^3 anaerobes/g of soil for all 18 samples. Plate counts for fungi showed the least variability of all the microflora, differing by no more than tenfold.

No identifications were performed for microfloral isolants from these soils. However, microscopic examinations showed that nearly all of the bacteria were bacilli. Many of them were chromogenic on agar plates, forming pink or yellow colonies. The fungi were represented by a few species of ascomycetous molds. Algae were generally a few species of filamentous oscillatoroid or nostocoid blue-greens. Some actinomycetes were also present, although much less abundant than the aerobic bacteria.

4. Conclusions

Numbers of aerobic, microaerophilic, and anaerobic bacteria as well as fungi and algae were present in 18 samples of soil taken from the surface 1 in of a high altitude field plot 24 ft on a side. Abundance of microflora in these samples was found to be microaerophilic bacteria > aerobic bacteria plus actinomycetes > anaerobic bacteria > algae > fungi. Relative variability of microflora from sample to sample was indicated as fungi < anerobic bacteria < aerobic bacteria plus actinomycetes < algae < microaerophilic bacteria. There was actually little variability in groups of microflora from sample to sample.

Additional effort is required to determine variations in abundance of groups of microflora in (1) other soil sites, (2) smaller and larger areas of similar soil, and (3) smaller as well as larger volumes of soil. Studies could also be undertaken to determine microbial variations between adjacent soil ecosystems of different genesis and correspondingly different properties, e.g., mineralogy, texture, salt, and organic matter content.

[11]Cameron, R. E., Morelli, F. A., and Blank, G. B., *Desert Soil Characteristics. I. Preliminary Studies*, Technical Report, Jet Propulsion Laboratory (to be published).

E. Fluorometric Determination of Urea. I. Determination of Urea with Diacetylmonoxime

J. Rho, J. R. Thompson, and J. V. Behar

1. Introduction

The need for a simple, yet accurate and reliable, method for the quantitative determination of urea in primate urine by an automated instrument for the Biosatellite primate testing program has resulted in the evaluation of a variety of direct and indirect procedures. The indirect methods depend upon the hydrolysis of urea with the enzyme urease, to form ammonia, with subsequent Nesslerization (Refs. 15–18). Several undesirable features of these methods are the use of relatively unstable reagents, loss of ammonia during the experimental process, and long analysis times.

The direct determination of urea depends upon the condensation of urea with some specific reagent such as α-isonitrosopropiophenone (Ref. 19) and diacetyl derivatives (Refs. 20–29) in acid media. The reaction with α-isonitrosopropiophenone requires special precautions because of the long heating time required for the production of color and because of the photosensitivity of the color formed. The methods using the reaction between urea and diacetyl or diacetyl derivatives appear to offer distinct advantages for the determination of urea. The various published modifications generally suffer from the fact that the color produced does not obey Beer's law.

The compound produced by the reaction of urea with diacetylmonoxime exhibits fluorescence (Ref. 30). The fluorescence maximum at a wavelength of 415 mμ with an optimal excitation wavelength of 380 mμ has been used for the quantitative determination of urea. However, the fluorescence intensity of the reaction products at this fluorescence maximum is not proportional to the concentration of urea.

In the present work the fluorescence properties of the reaction products of urea with diacetylmonoxime have been further studied with respect to the fluorescence maxima, the sample concentration, specificity, and kinetics.

2. Experiment

a. Apparatus. Fluorescence measurements were made on an Aminco-Bowman Spectrophotofluorometer using an Aminco 416-993 Xenon mercury lamp and an RCA IP28 photomultiplier detector.

b. Reagents.

Urea nitrogen stock standard. Urea nitrogen 1 mg per ml: 2.14 g of reagent grade urea were dissolved in 100 ml of 0.01-*N* sulfuric acid. Urea solution 10 m*M*: 0.6 g of urea was dissolved in 1.0 liter of ammonia-free water.

Diacetylmonoxime solution. 5.0 g of reagent grade diacetylmonoxime were dissolved in 500 ml of distilled water; 150 g of sodium chloride were added and the solution made up to 1 liter with water.

Urease. A 100-mg pellet of urease (Nutritional Biochemical Corp.) was dissolved in 50 ml of water containing 500 mg disodium ethylenediaminetetraacetate. The enzyme solution is made up fresh every month and kept refrigerated.

3. Procedure

The procedure adopted for the routine analysis of urea nitrogen is as follows: The urea sample is pipetted into a screw cap test tube (16 × 125 mm) and made up to 1.0 ml with 0.01-*N* sulfuric acid. To each tube, 0.8 ml of diacetylmonoxime and 0.2 ml of concentrated sulfuric acid are added. The tubes are capped tightly and placed in a boiling water bath for 20 min. At the end of 20 min the tubes are removed and allowed to cool to room temperature. A blank solution containing distilled water and standard solutions are run in the same manner. The fluorescence intensity is measured in the spectrophotofluorometer at wavelengths of 410 and 525 mμ with an excitation wavelength of 380 mμ. The concentration of the unknown urea solution is read from a standard curve prepared by treating urea solutions of known concentration in the same manner.

4. Results and Discussion

a. Reaction of urea with diacetylmonoxime. The reaction products of diacetylmonoxime and urea have an absorption maximum at 480 mμ as measured with a Cary spectrophotometer. This absorption band has been used by various authors for the determination of urea. The compounds produced by the reaction of urea with diacetylmonoxime also exhibit fluorescence at a wavelength of 410 mμ as observed by McCleskey (Ref. 30). However, the same reaction mixture fluoresces at another wavelength maximum of 525 mμ when the compound is activated at an optimal wavelength of 380 mμ. It was observed that the use of the second maximum of fluorescence is more advantageous for the determination of urea. The fluorescence intensities produced at the two different wavelength maxima were plotted against various concentrations of urea in Fig. 25. As shown in this figure, the use of the second fluorescence peak of the reaction product for the quantitative determination of urea has several advantages over the method using the first fluorescence peak.

The reaction of diacetylmonoxime with urea is not well understood. Natelson et al (Ref. 24) found the active reagent to be diacetyl, not diacetylmonoxime, whereas Hoseney and Finney (Ref. 31) found that diacetylmonoxime produced greater color intensity than diacetyl. The

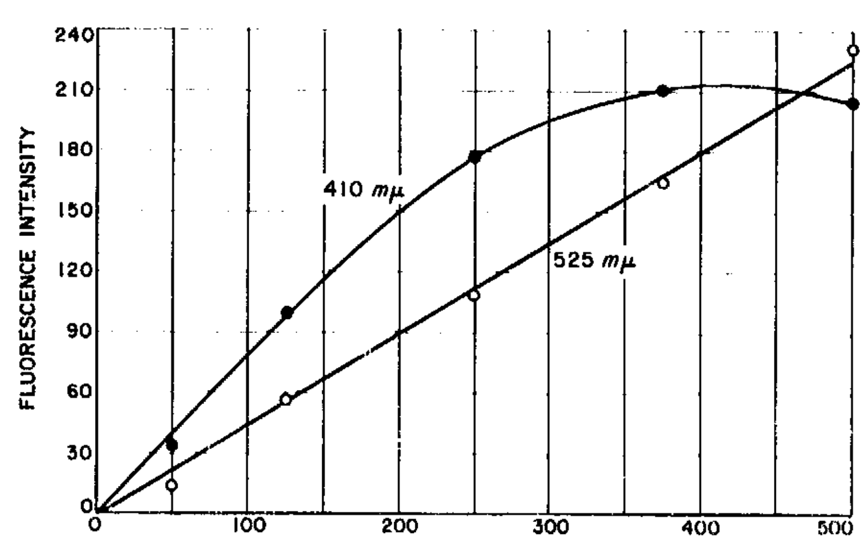


Fig. 25. Urea assay—comparison of standard curves for urea at 410 mμ and at 525 mμ

fluorescence maximum of the reaction products at 410 mμ can be observed when either diacetyl or diacetylmonoxime is used as the substrate, but the fluorescence maximum at 525 mμ is prominent only when a large excess of either reagent is used. The reaction products of urea with diacetylmonoxime exhibit the proportional increase of fluorescence intensity at 410 mμ at relatively low urea concentrations, but they quickly fail to maintain such proportionality as the concentration of urea is increased. On the other hand, the fluorescence intensity at 525 mμ shows proportionality with respect to sample size over a wide range of concentration. This is probably the result of different reaction products and also at least partially due to reduced interference by scattering of the incident light, since the 525-mμ peak is more remote from the exciting wavelength.

b. Specificity. The reaction with diacetylmonoxime apparently is positive for compounds containing the system RNHCO(S)N'R'', where R is hydrogen or a simple aliphatic radical; R' is hydrogen, a simple aliphatic radical, or a phenyl group; and R'' is less complex than a phenyl group (Refs. 20 and 31). The specificity of this reaction for urea has been studied by treatment of sample with and without the enzyme, urease. As shown in Fig. 26, the fluorescence intensities at both maxima increased substantially with various concentrations of samples which have not been subjected to the enzymic hydrolyses whereas the enzyme-treated samples failed to exhibit any sign of fluorescence at either maximum.

c. Effects of temperature. The production of fluorophores in the reaction varies with time at various temperatures. As shown in Fig. 27, the rate of formation of reaction products at 100°C is markedly higher than the rates at 60 and 80°C. When the reaction mixture was heated at 110°C, however, it was observed that the reaction was essentially complete at about 20 min.

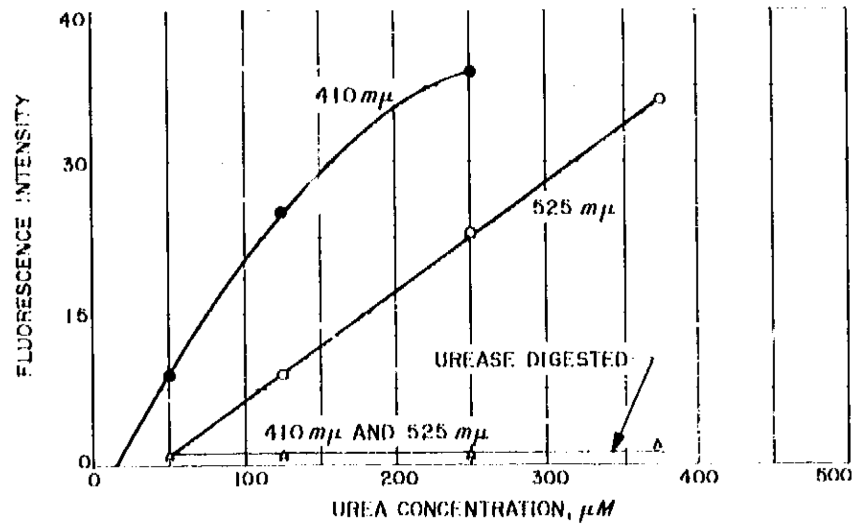


Fig. 26. Urea assay with and without urease treatment

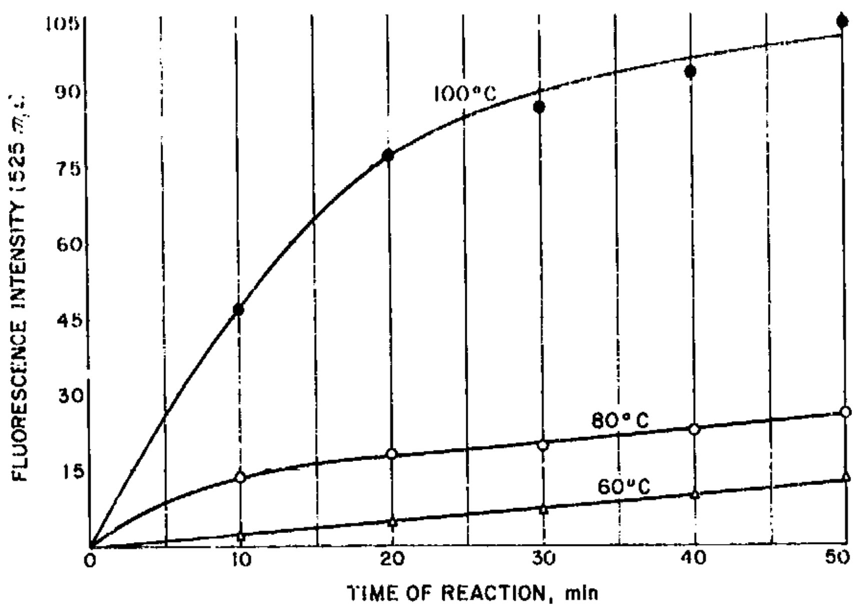


Fig. 27. Effect of temperature on urea assay

5. Summary

A method using diacetylmonoxime for the quantitative determination of urea has been modified and extended to determine urea. It has been observed that the products of the reaction of urea with diacetylmonoxime in acid solution exhibit two fluorescence maxima when irradiated at a single excitation wavelength; these are at 410 and 525 mμ. The use of the 525-mμ fluorescence peak of the reaction products for the quantitative determination has the advantage over the method using the first fluorescence peak in that the linearity of the fluorescence intensity with respect to the sample concentration extends over a greater portion of the fluorescence scale without decreasing the sensitivity of the method. The reaction is specific to urea as shown by the experimental results of urease digests.

F. Fluorometric Analysis of Urea. II. Determination of Urea in Urine Samples

J. Rho, J. R. Thompson, and J. V. Behar

1. Introduction

In order to study the physiological effects of prolonged weightlessness and accompanying stress by the *in situ* chemical analysis of the urine of a mammal flown in the

Biosatellite program, a simple and reliable method for estimating urea concentration is required. A direct single-step method is preferred for a small and completely automatic analytical laboratory in a flight experiment. Since the fluorometric method using diacetylmonoxime for the direct determination of urea appeared to be sufficiently accurate, it was applied for quantitative determination of urea in human and primate urine. In the present study, the reaction of diacetylmonoxime with urinary urea has been investigated with respect to the sample concentration, and its specificity and accuracy.

2. Experiment

a. Reagents. Reagents used in this study were the same as those described in the preceding article,[12] with the following additions.

Phenol color reagent. 5.0 g of phenol and 25 mg of sodium nitroprusside were dissolved in water and diluted to 500 ml with ammonia-free distilled water. The reagent was made fresh for every set of analyses.

Alkaline hypochlorite reagent. 2.5 g of sodium hydroxide and 0.21 g (1 ml) of sodium hypochlorite were dissolved in water and diluted to 500 ml with ammonia-free water. This reagent was also made fresh for every set of analyses.

A 40-m*M* ammonia stock standard was prepared by dissolving 2.641 g of ammonium sulfate in one liter of ammonia-free water. 10- and 20-m*M* solutions were prepared by an appropriate dilution of the 40-m*M* ammonia stock standard.

Hyland urine sample. Freeze-dried human urine samples were obtained from the Hyland Laboratories, Los Angeles 39, California. The samples were used after reconstituting the urine with water according to the directions given.

Primate (rhesus) urine samples. 24-hr pooled samples of primate urine were provided by Dr. Nello Pace of the University of California, Berkeley, California.

3. Procedure

The routine procedure followed for the fluorometric determination of urinary urea is the same as that described in the preceding article.[12] For the colorimetric determination of urinary ammonia and urea the method was adapted from that of Fawcett and Scott (Ref. 32) and Grunbaum[13].

Procedure for determination of ammonia. 10 μl each of a sample of 10-, 20-, and 40-m*M* ammonia working standard and urine were pipetted into separate test tubes (15×150 mm). 10 μl of ammonia-free distilled water were used as a blank. 5 ml of phenol color reagent were added to each tube and the contents were mixed. 5 ml of alkaline hypochlorite solution were immediately added to each tube and the contents were thoroughly mixed. The tubes were placed in a beaker of water at 60°C for 15 min and the optical density was measured at 630 mμ on the Cary Model 14 spectrophotometer.

Procedure for urea. 10 μl of the appropriately diluted urine samples were incubated with 50 μl of buffered urease at 37°C for 15 min along with an appropriate urea standard solution. After cooling to room temperature, the samples were treated as for the ammonia determination.

4. Results and Discussion

a. Specificity of the reaction of diacetylmonoxime with urinary urea. The reaction products produced by diacetylmonoxime with urinary urea have two fluorescence maxima, one at 410 mμ and the other at 525 mμ when the compounds are activated at an optimal wavelength of 380 mμ. The fluorescence intensity produced at these two maxima were plotted against urine concentration in Fig. 28. As is the case with the curves for pure urea, this figure shows that better linearity of the fluorescence intensity against the sample concentration is maintained with the fluorescence maximum at 525 mμ than with that at 410 mμ. Since it seems to be advantageous to use the fluorescence maximum at 525 mμ over that at 410 mμ, as discussed in the preceding article,[12] subsequent analyses were made using this emission wavelength.

The specificity of the reaction for urea was studied (1) by reacting diacetylmonoxime with normal urine and the specimens of urine from which urea was specifically removed by hydrolysis with urease and (2) by recovery of added urea from the urine sample. When various amounts of human urine were incubated with an appropriate amount of urease at 37°C for 15 min, the mixture was completely devoid of fluorescence, whereas un-

[12] "E. Fluorometric Determination of Urea. I. Determination of Urea with Diacetylmonoxime."

[13] Grunbaum, B. W., personal communication.

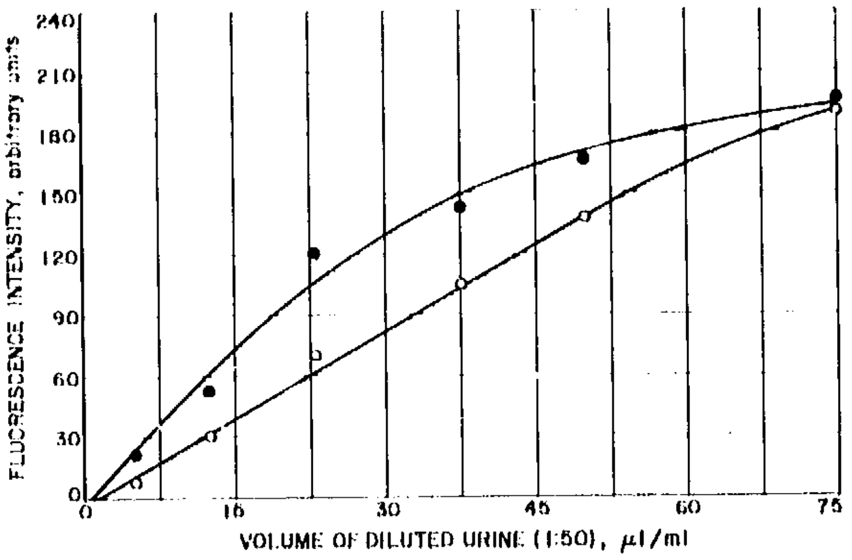


Fig. 28. Urinary urea assay

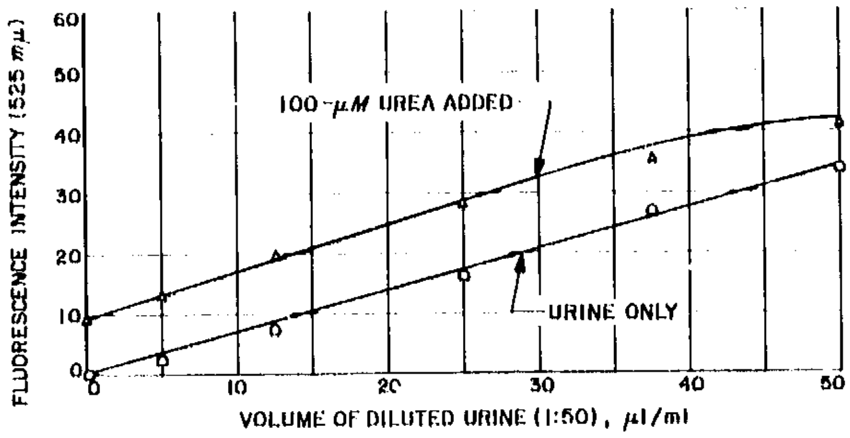


Fig. 30. Test for fluorescence recovery of urea in urine samples

treated urine exhibited the proportional increase of fluorescence with increased concentration of urine, as shown in Fig. 29. Recovery of urea was studied by adding aliquots of 100 μM urea to various concentrations of human urine and by measuring the fluorescence intensities of the reaction products of both urea-enriched and control samples. As shown in Fig. 30, equivalent increments of the fluorescence intensity is observed throughout the concentration range of urine samples used. Thus, the output of fluorescence is not only specific to the urea present but also proportional to the concentration of urea in the samples.

b. Quantitative fluorometric determination of urinary urea. The concentration of urea in the Hyland Laboratory and primate urine determined by this procedure is shown in Table 9. Mean values of 0.207 m mole per ml for the Hyland urine and 0.122 m mole per ml for the primate urine were obtained. This value for urea in the Hyland urine agrees very closely with their data of 6 mg of urea nitrogen per ml. Analytical data given by Hyland Laboratories were derived by optical density measurements following the Nesslerization of urease-treated samples. Excellent agreement was also obtained for the primate urine between the fluorometric data and the value derived from the colorimetric assay employing urease digests[14] (Ref. 32). The value 0.12 m mole of urea per ml of the same batch of primate urine was independently obtained by Grunbaum at another laboratory[14].

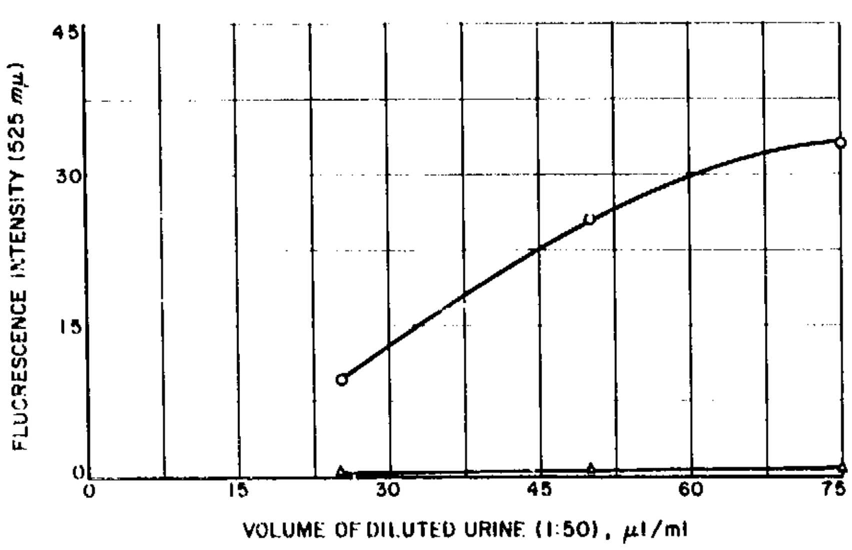


Fig. 29. Urinary urea assay with and without urease treatment

Table 9. Fluorometric measurement of urea in urine samples

Samples	Fluorescence intensity at 525 mμ			Urea concentration, m mole/ml
	Assay 1	Assay 2	Mean value	
Urea standard				
100 μM	8.5	9.3	8.9	0.100
125 μM	10.5	11.5	11.0	0.125
175 μM	18.8	20.5	19.6	0.175
250 μM	29.0	29.0	29.0	0.250
375 μM	41.8	43.5	42.0	0.375
500 μM	48.5	57.5	53.0	0.500
Hyland urine				
0.75 μl/ml	15.8	17.8	16.8	0.210
1.00 μl/ml	23.0	25.5	24.2	0.210
1.25 μl/ml	29.8	31.0	30.4	0.208
1.75 μl/ml	39.2	41.7	40.4	0.200
Primate urine				
1.75 μl/ml	24.5	25.5	25.0	0.123
2.50 μl/ml	37.5	32.5	35.0	0.120

[14] Grunbaum, B. W., personal communication.

5. Summary

The reaction of diacetylmonoxime with urea has been applied to determination of urea in human and primate urine. The reaction is specific to urinary urea as shown by the experimental results of urease digests and by the recovery study of added urea in urine samples. Comparative study with colorimetric determination of ammonia following urease treatment of urine shows the method to be valid.

G. A New Technique for the Enzymatic Detection of Hydrogen Peroxide

H. H. Weetall and N. Weliky

The several available methods for the detection and estimation of hydrogen peroxide may be divided into three main categories: (1) direct measurement of the absorbance of the hydrogen peroxide ultraviolet spectrum, (2) oxidation or reduction of hydrogen peroxide by some compound to yield a detectable product, and (3) detection of the product of an enzymatically catalyzed reaction. The first two techniques can be used satisfactorily for hydrogen peroxide concentrations as low as 10^{-3} M if interfering substances are absent and sufficient quantities of solution are available. Enzymatic methods can detect 10^{-6} M peroxide.

Enzymatic methods for detecting the presence of small quantities of peroxides have been used since the first successful attempt in 1845 by Osann (Ref. 33). With peroxidase as the enzyme he was able to detect 6×10^{-5} M hydrogen peroxide. The method made use of peroxidase to catalyze the oxidation of a colorless hydrogen donor (DH_2) to a colored product (D) which could be determined colorimetrically (Ref. 33). Schematically the reaction is as follows:

$$H_2O_2 + DH_2 \xrightarrow{\text{peroxidase}} 2H_2O + D$$

Typical donors are organic amines, phenols, and quinones. Solutions must be made fresh daily.

We now have developed a rapid semi-quantitative spot test for hydrogen peroxide which can be used to detect peroxide in small volumes at concentrations as low as 10^{-6} M. The method utilized an enzyme-paper preparation which still retains activity after being stored for over two months without refrigeration. The reagent paper was prepared by chemically coupling horseradish peroxidase (Worthington Biochemical Corporation) to carboxymethylcellulose paper strips (0.7 meq/g) in the presence of N,N′-dicylohexylcarbodiimide (Ref. 34). The resulting peroxidase-paper was washed exhaustively with distilled water and dried at room temperature between paper blotters.

Various concentrations of hydrogen peroxide ranging from 1.15 M to 1.15×10^{-6} M were prepared in distilled water. An aliquot of each concentration was diluted with an equal volume of a saturated solution of benzidine in 0.1-M tris buffer, pH 8.5. Samples of 0.02 ml spotted on a strip of peroxidase-paper immediately turned blue, then slowly became brown. The intensity of the color was proportional to the hydrogen peroxide concentration used. In a control experiment, in which the peroxide was omitted, no immediate blue color appeared but within 30 min the spot had turned brown due to air oxidation. Fig. 31 shows the difference in intensity of spots made with different concentrations of peroxide. The photograph was taken about an hour after the solutions were applied to the paper. The intensities at very low concentrations are not differentiated well by black and white photography.

Different enzymes and reagents will be coupled to investigate the possibility of detecting and estimating other low molecular weight inorganic and organic molecules of interest, and also biological macromolecules.

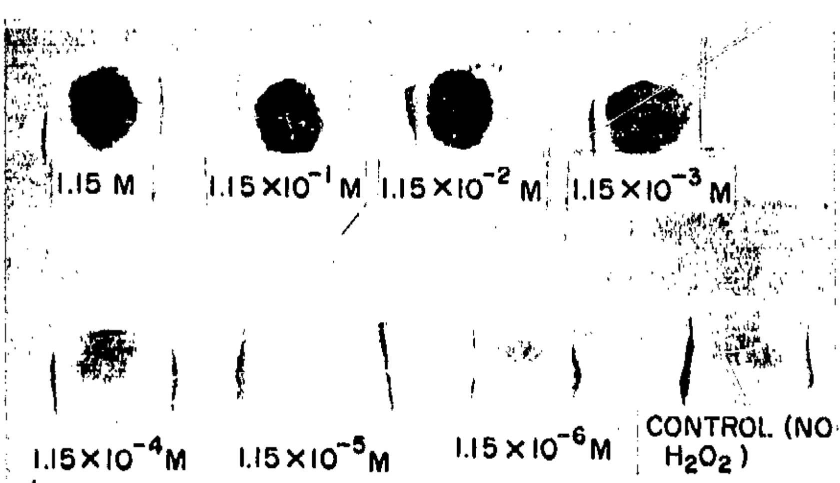


Fig. 31. Decreasing intensity of spots on CMC-peroxidase paper with decreasing concentration of hydrogen peroxide

References

1. Eskind, N., "Sample Collection for Mars Biological Experiments," *Space Programs Summary* No. 37-23, Vol. IV, pp. 275–278, Jet Propulsion Laboratory, Pasadena, California, October 31, 1963.

2. Cameron, R. E., *The Role of Soil Science in Space Exploration*, Technical Report No. 32-399, Jet Propulsion Laboratory, Pasadena, California, February 16, 1963.

3. Griggs, R. E., "The Valley of Ten Thousand Smokes," The National Geographic Society, Washington, D.C., 340 pp., 1922.

4. Smith, N. R., and Griggs, R. F., "The Microflora of the Ash of Katmai Volcano with Especial Reference to Nitrogen Fixing Bacteria," *Soil Science 34*, pp. 365–373, 1932.

5. Cahalane, V. H., *A Biological Survey of Katmai National Monument*, Smithsonian Miscellaneous Collections, 38 (5), pp. 1–246, 1959.

6. Cameron, R. E., and Blank, G. B., *Soil Organic Matter*, Technical Report No. 32-443, Jet Propulsion Laboratory, Pasadena, California, May 23, 1963.

7. Bollen, W. B., *Microorganism Study-Culture and Identification*, Contract No. 950783, Oregon State University, Corvallis, Oregon.

8. Williams, H., Curtis, G., and Juhle, W., "Mount Katmai and the Valley of Ten Thousand Smokes, Alaska," *VIII Pacific Science Congress*, 2:129, 1953.

9. Quimby, F., Editor, *Concepts for Detection of Extraterrestrial Life*, NASA SP-56, 53 pp., 1964.

10. Heim, A., "Radioisotopic Metabolic Detection of Possible Martian Lifeforms," *Proceedings of the Lunar and Planetary Exploration Colloquium*, Vol. 3, pp. 37–46, 1963.

11. Vishniac, W., *Development of a Life Detector for Planetary Soils*, NASA Grant NsG-19-59, 7 pp., 1961.

12. Hallam, M. J., and Bartholomew, W. V., "Influence of Rate of Plant Residue Addition in Accelerating the Decomposition of Soil Organic Matter," *Soil Science Society of America Proceedings*, Vol. 17, pp. 365–368, 1953.

13. Stuart, J., "Instrumentation Requirements for Life Detection Systems," *Bioengineering Symposium*, pp. 88–97, U.S. Air Force Academy, May 4–5, 1964.

14. Russell, E. W., "Soil Conditions and Plant Growth, Ninth Edition," pp. 135–168, John Wiley & Sons, Inc., New York.

15. Looney, J. M., *Journal of Biological Chemistry*, Vol. 88, p. 189, 1930.

16. Hughes, J., and Saifer, A., *Journal of Laboratory and Clinical Medicine*, Vol. 27, p. 381, 1941.

17. Day, H. G., Bernstorf, E., and Hill, R. T., *Analytical Chemistry*, Vol. 21, p. 1290, 1949.

References (Cont'd)

18. Kibrick, A. C., and Skupp, S., *Proceedings of the Society for Experimental Biology and Medicine*, Vol. 73, p. 432, 1950.

19. Archibald, R. M., *Journal of Biological Chemistry*, Vol. 157, p. 507, 1945.

20. Fearon, W. R., *Biochemical Journal*, Vol. 33, p. 902, 1938.

21. Ormsby, A. A., *Journal of Biological Chemistry*, Vol. 146, p. 595, 1942.

22. Barker, S. B., *Journal of Biological Chemistry*, Vol. 153, p. 453, 1944.

23. Wheatley, V. R., *Biochemical Journal*, Vol. 43, p. 420, 1948.

24. Natelson, S., Scott, M. L., and Beffa, C., *American Journal of Clinical Pathology*, Vol. 21, p. 275, 1951.

25. Friedman, H. S., *Analytical Chemistry*, Vol. 25, p. 662, 1953.

26. Dickenman, R. C., Crafts, B., Zak, B., *American Journal of Clinical Pathology*, Vol. 34, p. 981, 1954.

27. Rosenthal, H. L., *Analytical Chemistry*, Vol. 27, p. 1980, 1955.

28. Lemar, R. L., and Bootzin, D., *Analytical Chemistry*, Vol. 29, p. 1233, 1957.

29. Richter, H. J., and Lapointe, Y. S., *Clinical Chemistry*, Vol. 5, p. 617, 1959.

30. McCleskey, J. E., *Analytical Chemistry*, Vol. 36, p. 1647, 1964.

31. Hoseney, R. C., and Finney, K. F., *Analytical Chemistry*, Vol. 36, p. 2145, 1964.

32. Fawcett, J. K., and Scott, J. E., *Journal of Clinical Pathology*, Vol. 13, p. 156, 1960.

33. Bergmeyer, Hans-Ulrich, *Methods of Enzymatic Analysis*, Academic Press, New York, 1963.

34. Weliky, N., Weetall, H. H., Gilden, R. V., and Campbell, D. H., *Immunochemistry*, Vol. 1, p. 219, 1964.

XX. Fluid Physics

A. Inviscid Eigenvalues and the Reynolds Number Dependence of the Eigenvalues of the Complete Laminar Stability Equations

L. M. Mack

In Ref. 1, two separate families of solutions of the complete stability equations at finite Reynolds numbers were found for a highly cooled boundary-layer solution. In the present contribution, the eigenvalues of these multiple solutions are studied further and are related to the eigenvalues of the inviscid solutions. From this study it is possible to demonstrate that the eigenvalues of the inviscid equations are indeed the infinite Reynolds number limit of the eigenvalues at finite Reynolds number obtained from the complete stability equations. For this demonstration it is necessary to know all of the inviscid solutions that exist for a given wave number, α, and not just the neutral and amplified solutions which have been considered in previous *Space Programs Summaries*.

The eigenvalue diagrams for the insulated-wall boundary layer at $M_1 = 4.2$ are given in Fig. 1. For the range of wave numbers in the figure, there are only four neutral solutions, the sonic neutral solution with the eigenvalues $\alpha = 0$, $c_r = 1 - 1/M_1$ (c_r is the phase velocity), the first- and second-mode neutral solutions, α_{s1}, c_s, and α_{s2}, c_s, associated with the generalized inflection point, and the second-mode singular neutral solution, α_{12}, 1. The curve given by the eigenvalues of the first-mode amplified solutions is seen to be continued beyond α_{s1}, c_s by the eigenvalues of a family of damped solutions. These solutions, which initially have the first-mode characteristic of little phase change across the boundary layer in the pressure fluctuation, take on the second-mode feature of a marked phase change in the region of maximum negative c_i. It is of interest to note that the maximum rate of damping and the maximum rate of amplification both occur at almost the same value of α. The curve given by the eigenvalues of the second-mode amplified solutions is continued beyond α_{s2}, c_s by the eigenvalues of a family of lightly damped solutions. For these solutions, the phase velocity c_r decreases rapidly with increasing α, and eventually the solutions become supersonic, i.e., $c_r < 1 - 1/M_1$, and the waves travel at a supersonic speed relative to the free stream.

As M_1 increases, the appearance of the eigenvalue diagrams changes. For instance, it is found that the eigenvalues of the supersonic damped solutions are no longer reached by following the curve through the eigenvalues of the first-mode damped solutions, but by following the

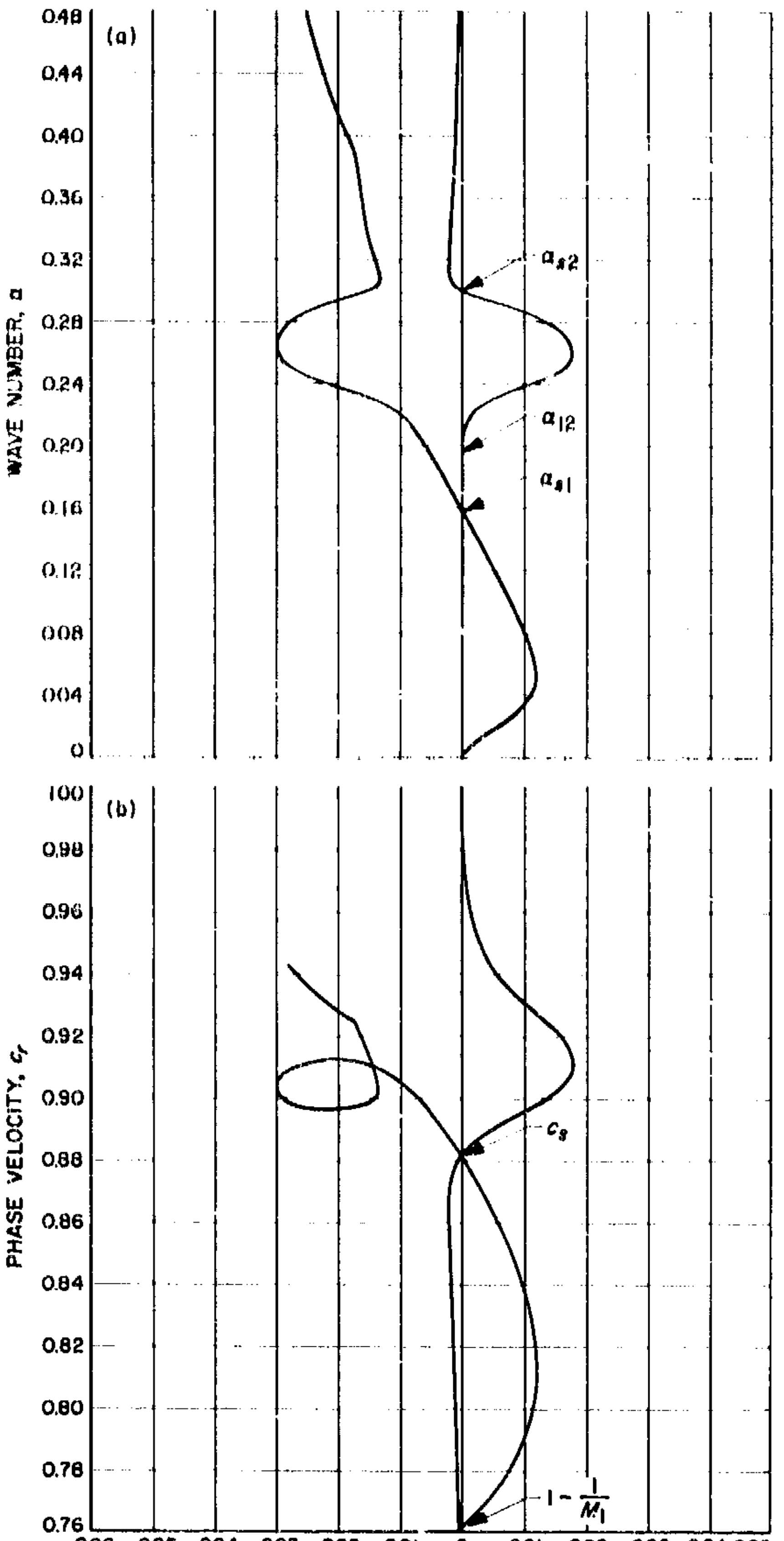


Fig. 1. Inviscid eigenvalues for insulated-wall boundary layer at $M_1 = 4.2$; (a) α vs c_i, (b) c_r vs c_i

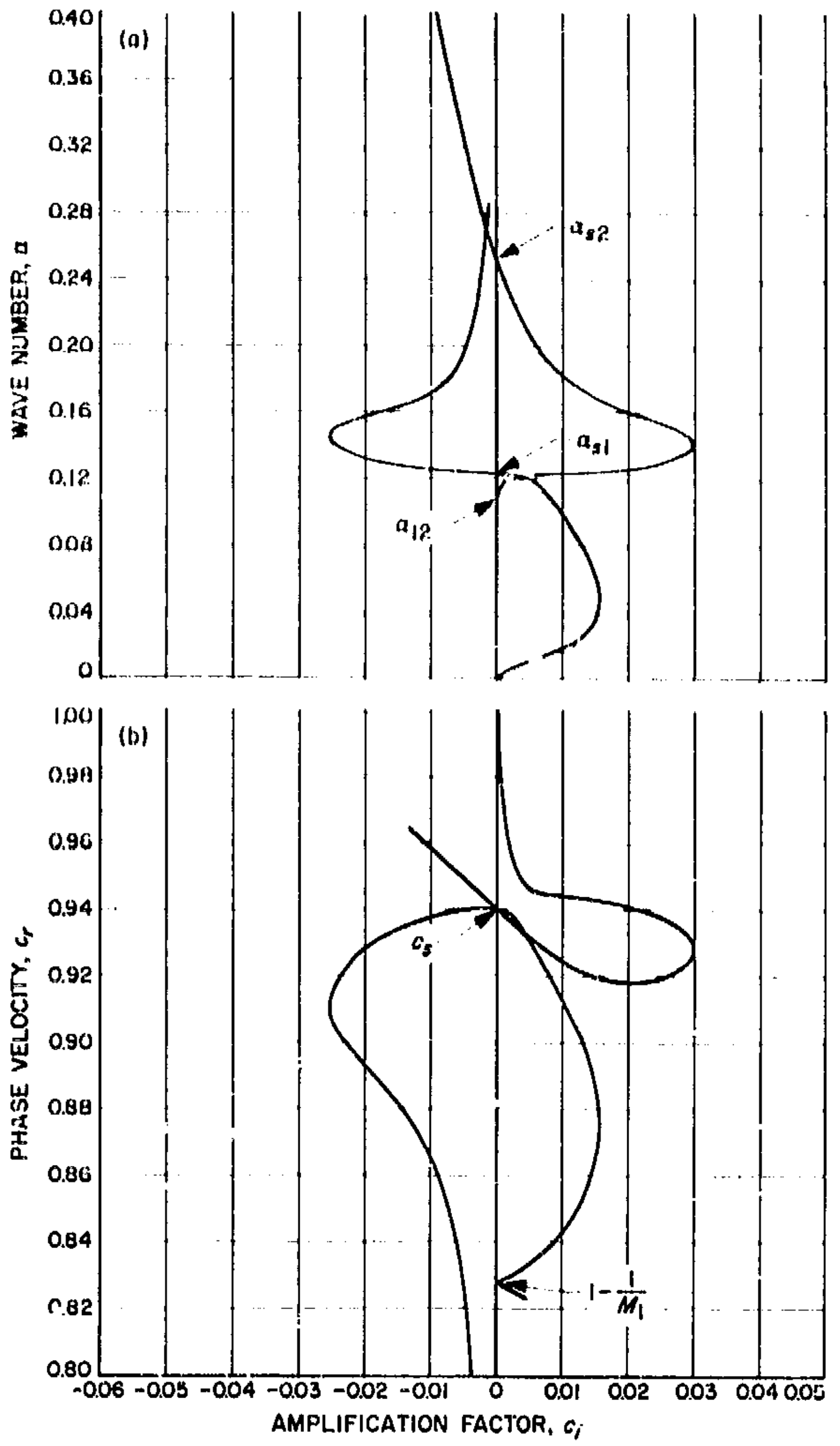


Fig. 2. Inviscid eigenvalues for insulated-wall boundary layer at $M_1 = 5.8$; (a) α vs c_i, (b) c_r vs c_i

curve through the eigenvalues of the second-mode amplified solutions. This change can be seen in the diagrams for the insulated-wall boundary layer at $M_1 = 5.8$ shown in Fig. 2. Cooling the boundary-layer wall moves α_{12} away from α_{s1}, and in this respect returns to the situation found at lower Mach numbers for the insulated-wall boundary layer. Therefore, sufficient cooling diagrams of the Fig. 1 type are again obtained, but with the difference that for the cooled-wall boundary layer the first-mode is stabilized. Fig. 3 (solid lines) gives eigenvalue diagrams of this kind for $M_1 = 5.8$, $T_w/T_r = 0.25$ (T_w is the wall temperature, T_r is the insulated-wall temperature).

In all of these eigenvalue diagrams there are at least two inviscid solutions for $\alpha \geq \alpha_{12}$. For $\alpha \geq \alpha_{13}$, there will be at least three solutions, etc. A natural question to ask

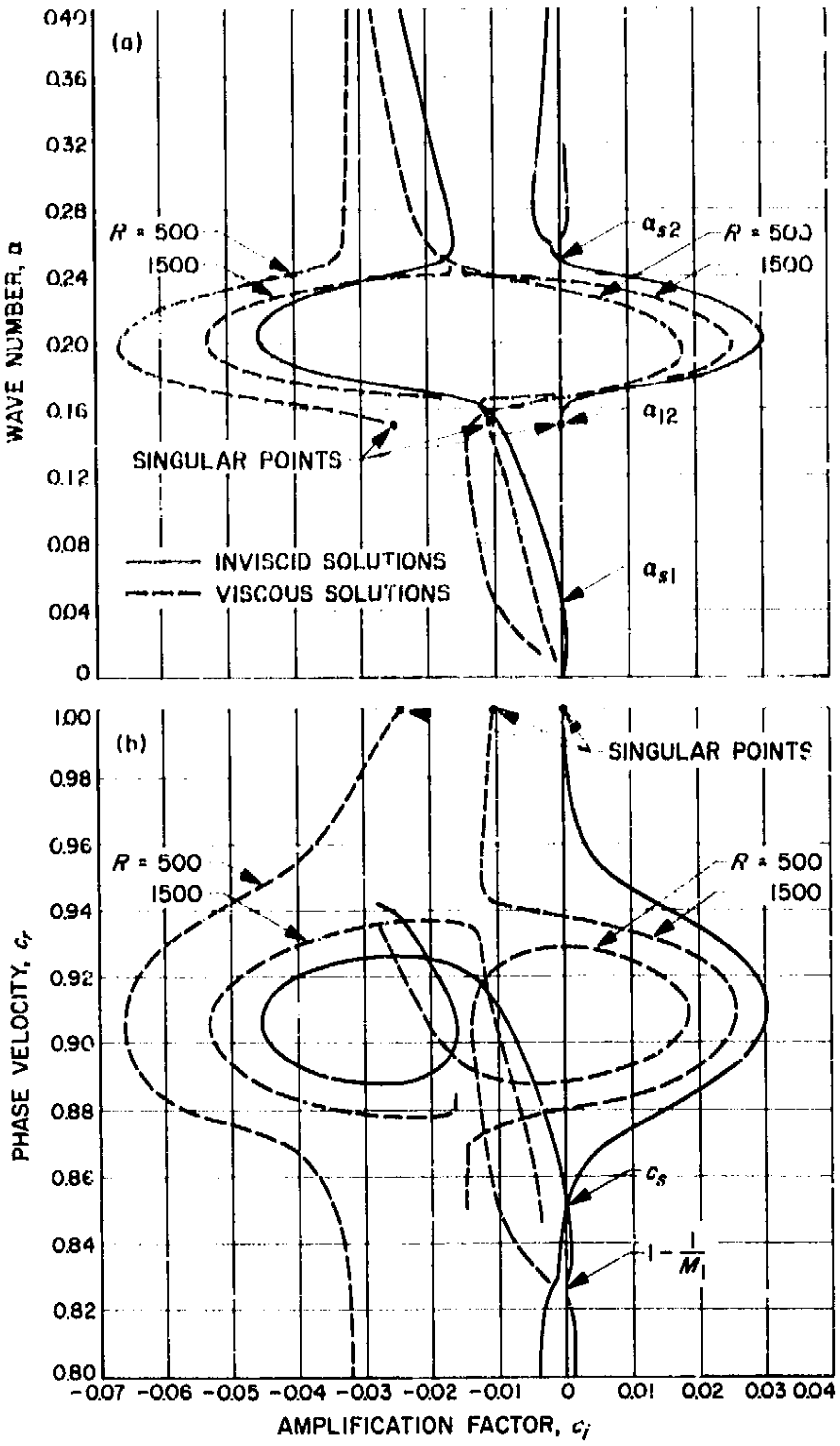


Fig. 3. Viscous and inviscid eigenvalues for cooled-wall boundary layer at M_1 = 5.8, T_w/T_r = 0.25; (a) α vs c_i, (b) c_r vs c_i

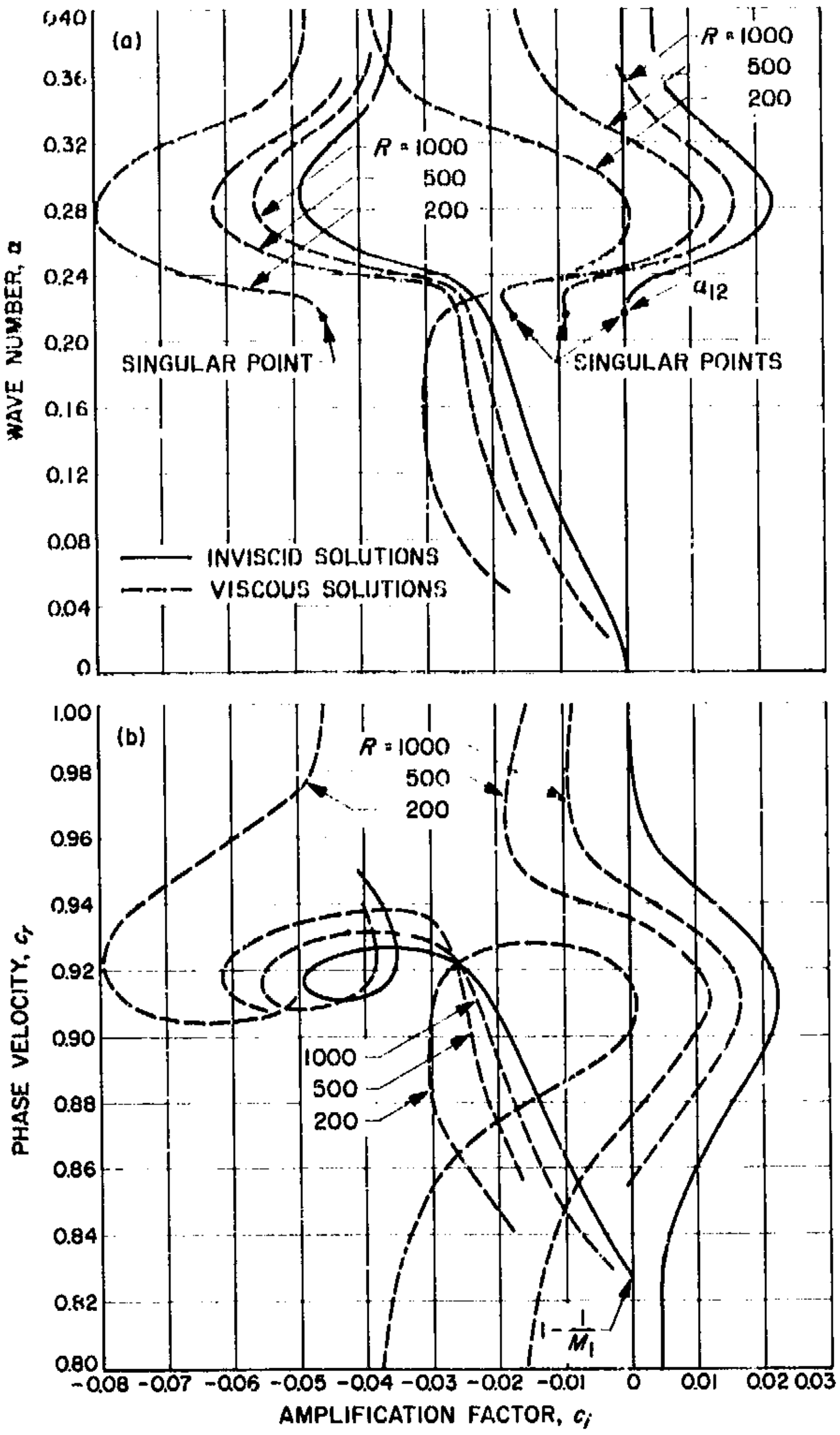


Fig. 4. Viscous and inviscid eigenvalues for cooled-wall boundary layers at M_1 = 5.8, T_w/T_r = 0.05; (a) α vs c_i, (b) c_r vs c_i

at this point is whether each of these solutions corresponds to a separate solution at finite Reynolds numbers, or whether the introduction of a non-zero viscosity makes the solution unique.

Originally, it was thought that the second alternative was the correct one because computations of viscous solutions at a fixed Reynolds number and over a wide range of wave numbers had never indicated the presence of a second viscous solution. However, when a constant-frequency disturbance was extended to lower Reynolds numbers in the damped region for $M_1 = 5.8$, $T_w/T_r = 0.05$, it was rather surprisingly found (Ref. 1) that the disturbance did not extend back toward $R = 0$ but instead led to a counterpart of the inviscid singular neutral solution located at a Reynolds number considerably greater than zero.

The eigenvalues for the above case are given in Fig. 4 for both the inviscid solutions (solid lines) and the finite-Reynolds number solutions (dashed lines). The boundary layer of Fig. 4 has no generalized inflection point and the first mode is completely stable. However, there is still a family of inviscid solutions which starts at the sonic neutral solution, but which consists only of damped solutions. These solutions have the character of first-mode solutions below the region of maximum damping. At $R = 1000$, the highest Reynolds number at which amplified solutions can be computed, there are two viscous solutions corresponding to the two inviscid solutions. The eigenvalue diagrams of the viscous solutions are of exactly the same type as for the inviscid solutions. The only difference is that the stabilizing effect of viscosity has shifted the curves for $R = 1000$ to the left of the inviscid curves, i.e., to smaller values of c_i. It is the family of solutions whose eigenvalues lie on the curve extending from the singular point that includes the amplified solutions.

At $R = 500$, the eigenvalues follow the same pattern as at $R = 1000$, but with the curves moved even further to the left. However, at $R = 200$, the singular point has crossed to the left of the curve coming from $\alpha = 0$, and there is a distinct change in the arrangement of the curves. At this Reynolds number it is the curve starting at $\alpha = 0$ that leads to the amplified solutions; the family of solutions that includes the singular point contains only damped solutions. It is apparent that if for $R = 1000$, a frequency corresponding to one of the amplified solutions is extended back to smaller Reynolds numbers, the singular point will be reached rather than $\alpha = 0$, $R = 0$, provided the singular point is at a Reynolds number where the eigenvalue diagrams of the two solutions are still the same as at $R = 500$ and 1000. However, if the same calculation is performed at a low Reynolds number where the amplified solutions lie on the other curve, then the point $\alpha = 0$, $R = 0$ will be reached. One further change not shown in Fig. 4 occurs at still lower Reynolds numbers. At $R = 100$, it is the curve that starts at the singular point rather than at $\alpha = 0$ that leads to the supersonic damped solutions. The latter arrangement is thus the same as at $R = 500$ and 1000, but with the position of the two curves reversed in the diagrams.

The eigenvalue diagrams at $R = 500$ for $M_1 = 5.8$, $T_w/T_r = 0.25$ are shown by the dashed lines in Fig. 3. This type of diagram, which is of the same type as found at $R = 100$ for $T_w/T_r = 0.05$, persists to at least $R = 1400$. At $R = 1500$, the diagrams are the same as for the inviscid solutions. If the intermediate arrangement found at $R = 200$ for $T_w/T_r = 0.05$ exists in this case, it does so only for a very small range of Reynolds numbers. For still higher wall temperatures, the type of eigenvalue diagrams corresponding to the inviscid solutions cannot be reached within the Reynolds number range of the computation. The solutions on the curve which starts at the singular point are all damped and correspond to the damped portion of the curve for the inviscid solutions which starts at $\alpha = 0$.

The main conclusion is clear. No matter what the arrangement of the eigenvalues at low Reynolds number, for sufficiently large Reynolds numbers the eigenvalue diagrams are of the same type as for the inviscid solutions. This result demonstrates that the inviscid solutions are indeed the infinite Reynolds number limit of the viscous solutions.

The lowest Reynolds number at which the inviscid type of eigenvalue diagrams is found appears to depend on the separation between α_{12} and α_{s1}. The greater the difference between α_{12} and α_{s1}, the lower the Reynolds number at which the inviscid arrangement is found. Consequently, it can be expected that for the insulated-wall boundary layer at much lower Mach numbers than 5.8, the second mode amplified solutions will originate from the singular point even at fairly low Reynolds numbers. As a result, a constant-frequency disturbance starting near $R = 0$ will not lead to these solutions, but only to the first-mode amplified solutions. The constant-frequency disturbance that leads to the second-mode amplified solutions will originate at the singular point for that frequency located at some non-zero Reynolds number.

A rather striking relationship between the viscous and inviscid eigenvalues is found when c_r is plotted against α for both the viscous and inviscid solutions. Such a plot is shown in Fig. 5 for $M_1 = 5.8$, $T_w/T_r = 0.25$. The dashed lines refer to the viscous solution at $R = 500$, and the solid lines to the inviscid solution. Third-mode eigenvalues are also shown in this figure. The close correspondence of the viscous and inviscid eigenvalues is remarkable when it is considered that the Reynolds number of the viscous eigenvalues is low, and that the values of c_i are completely different on the two curves (see Fig. 3). For the inviscid curve that starts at $\alpha = 0$, only the portion up to α_{s1} has positive c_i. The solutions on the remainder of the curve are all damped. The inviscid second-mode amplified solutions are on the curve that starts at α_{12}. On the contrary, the viscous curve that starts near α_{12} includes only damped solutions. The viscous second-mode amplified solutions are on the other curve, and are located between the indicated neutral points.

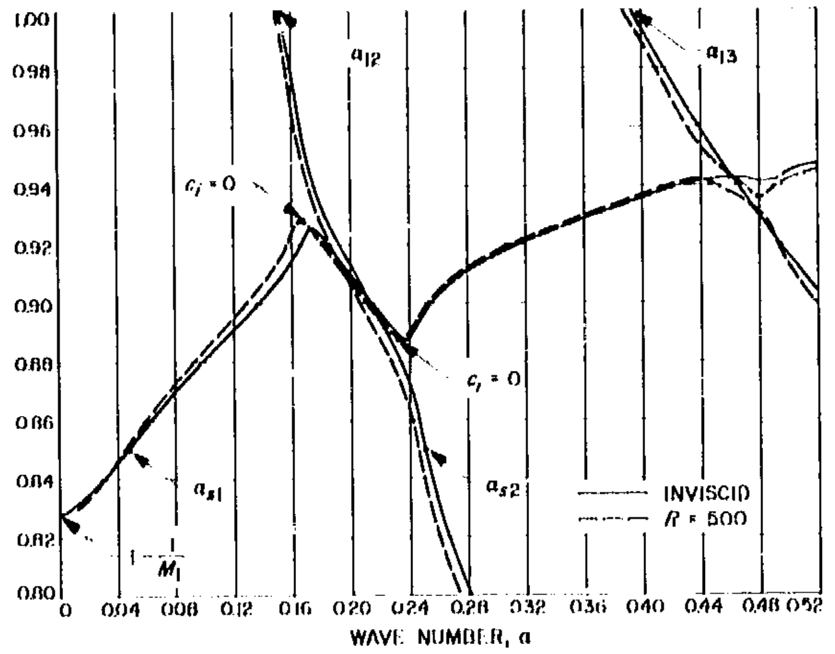


Fig. 5. Phase velocity vs wave number for cooled-wall boundary layer at $M_1 = 5.8$, $T_w/T_r = 0.25$

For $R = 1500$, the appearance of the two viscous curves (for wave numbers below the third-mode region) is much the same as for $R = 500$. However, as previously mentioned, the amplified solutions are found on the curve starting at the singular point rather than on the one starting at $\alpha = 0$. This change is obviously necessary if the inviscid solutions are to be the $R \rightarrow \infty$ limit of the viscous solutions.

The arrangement of the viscous curves in the third-mode region in Fig. 5 is one that can also occur in the second-mode region. For example, it occurs for $R = 200$ at $M_1 = 5.8$, $T_w/T_r = 0.05$ (Fig. 4), and also for the inviscid eigenvalues at $M_1 = 5.8$, insulated wall (Fig. 2). Again, the chief factor in determining the arrangement concerns which curve leads to the supersonic damped solutions.

An additional point that can be discussed with reference to Fig. 5 is the group velocity. It will be recalled that in order to transform the time rate of amplification of the theory into the space rate of amplification needed to compute the over-all amplification of fixed frequency disturbances, a wave propagation velocity is required. The phase velocity c_r was used for this purpose in Ref. 1, although it is more customary to use the group velocity. The latter is defined by

$$c_g = c_r + \alpha \frac{dc_r}{d\alpha}$$

It is apparent from Fig. 5 that near the two extrema of the $R = 500$ curve extending from $\alpha = 0$, the group velocity changes almost discontinuously by about 25%. It is most unlikely that the correct propagation velocity for the time-space transformation should undergo such a change. Fig. 5 is suggestive of a curve of index of refraction vs frequency for a dispersive medium in the theory of electromagnetic waves. The range of wave numbers for which $dc_r/d\alpha$ changes sign is thus similar to a region of anomalous dispersion (with the sign of $dc_r/d\alpha$ in the normal region the opposite of what it is for an electromagnetic wave). In a region of anomalous dispersion, the velocity of energy propagation is neither the group velocity nor the phase velocity, but a third velocity called the signal velocity. This velocity, which is identical to the group velocity in a region of normal dispersion, is closer to the phase velocity than to the group velocity near a resonant frequency.

B. Spectroscopic Measurement Near a Plasma Shock Wave

W. H. Christiansen

Electron number density and temperature profiles are being measured in front of a flat disc in a supersonic seeded-gas plasma. The region of interest extends several body diameters upstream of the disc. The technique of measurement and initial experimental results are reported here.

The experiment is being carried out in the low-density wind tunnel. Thermally ionized cesium is seeded into argon (Ref. 2) to produce the plasma. The seed gas density controls the electron density. At typical stagnation conditions, the maximum cesium concentration by mole is 0.5% that of argon or $4(10^{14})$ atoms/cm^3. About 25% of the cesium is ionized.

A 2.5-cm disc is held normal to a flow with a Mach number of 5.2. At this Mach number, the free-stream mean free path of argon is 0.5 mm with a corresponding shock thickness of 1.5 mm. The normal shock standoff distance is 8 mm while the boundary layer is approximately 2 mm thick. Hence, under these conditions the flow regions are not merged.

The flow is analyzed spectroscopically using a Jarrell-Ash ½-m spectrometer with fixed slits. For the measurements, the exit slit was made 5 times larger than the entrance slit. A photomultiplier tube detected the output of the spectrometer. The image of the flow region of interest is focused onto the entrance slit with a 12.8-cm achromatic lens set at twice the focal length from the object. This lens, together with the slit dimensions of 0.0040×0.25 cm, gives a resolution of 0.125×0.375 cm for a 2.5-cm-diameter plasma flow, which is more than sufficient to accurately measure the distribution of light intensity everywhere except in the shock wave itself. A movable mirror placed in the optical path of the lens provides for scanning of the image along the jet as well as across the jet (see Fig. 6). This mirror has electrical readouts so that the position of the image relative to the spectrometer is known. The entire optical system, as used in the wind tunnel, was calibrated absolutely with a tungsten lamp of known spectral intensity. The wavelength of interest extends from 4000 to 8000Å.

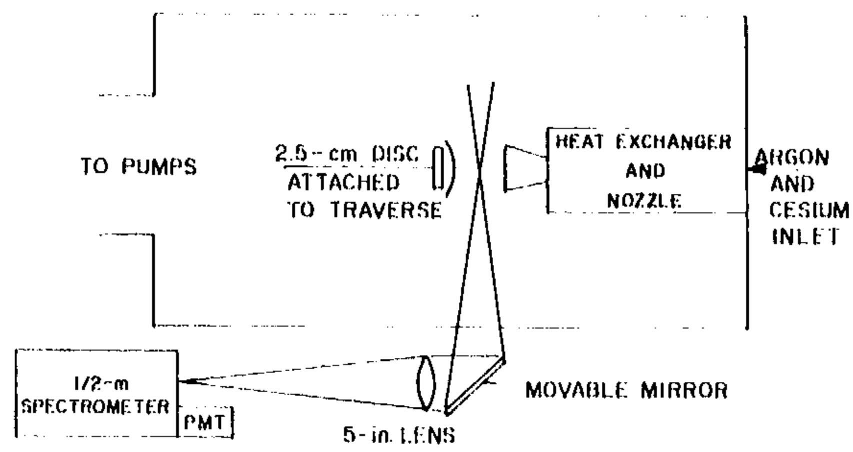


Fig. 6. Schematic of experimental setup

The method of measurement was as follows. Absolute line intensity measurements were taken of the spectrum of cesium in the jet without the disc. Using the theoretical oscillator strengths of Stone (Ref. 3), the log of the number density of excited atoms (n) divided by the multiplicity of the state (g) was plotted versus the energy level of the atom. A straight line portion on this plot (Boltzmann plot) indicates a form of thermal equilibrium between the important collisions. It is known that the important collisions are electron-atom collisions at high energy levels in the atom. A typical result is shown in Fig. 7. The results fall on a straight line at energy levels in excess of 3.5 ev. It was felt that these lines do indicate an equilibrium between electrons and atoms of cesium. From this result it was decided that it was necessary to measure only the violet components of the diffuse series with energy levels of the atom in excess of 3.5 ev.

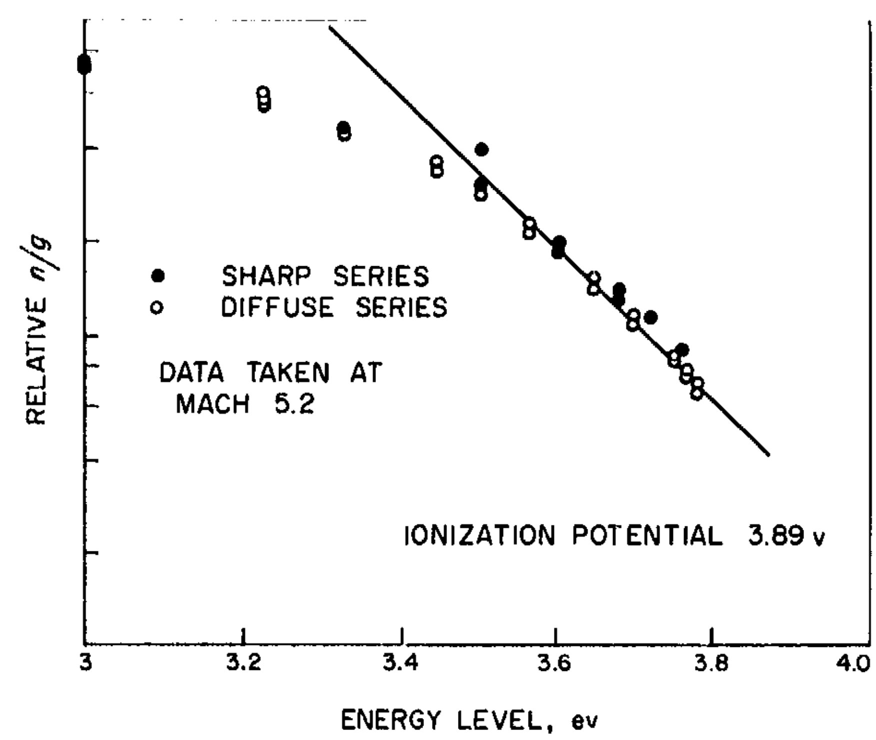


Fig. 7. Relative density of the excited states of cesium divided by the multiplicity of the state vs the binding energy of the state

The disc is inserted in the flow downstream of the point of observation. The technique is to scan across the jet at a fixed axial position for the intensity measurement of a spectral line and then repeat for as many lines as required. Then the disc is moved closer to the point of observation and the measurements are repeated. This is continued until a flow region of several body diameters upstream of the disc has been investigated. These measurements of intensity versus distance across the jet are converted to intensity as a function of radius through Abel's inversion. The centerline results are then plotted. The slope of the Boltzmann plot gives electron temperature, while the intensity level gives electron density.

The first results of this measurement indicate that the electron temperature is constant in the region of the shock, but the number density varies. In particular, the electron density seems to decrease by as much as 50% as the shock wave is approached. This gives a dark space in front of the very luminous shock layer. At the shock wave itself, the electron density exhibits a density rise of about four.

At present no explanation of this result can be found. A theoretical solution (Ref. 4) of the shock wave in a plasma indicates that electron temperature should vary in front of the shock, but the number density should not. This, of course, is opposite that of the measurements. Because cesium is heavier than argon, relative diffusion or separation by cesium from argon may take place in the shock wave (Ref. 5). However, the dark space extends for many centimeters upstream of the shock wave, much greater than the shock wave thickness. Further measurements on the phenomena are being carried out. Utilizing a very low-noise photomultiplier tube, Stark broadening measurements will be attempted. If these results show a similar trend, a systematic change of variables will be carried out to get the important variations of this phenomena.

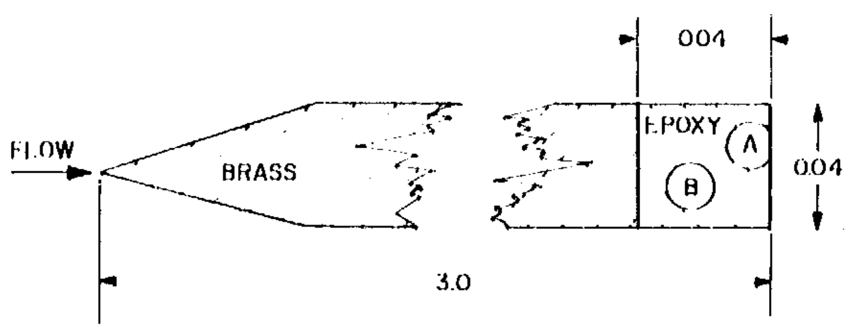


Fig. 8. Wind tunnel model with wake disturbance generator

C. A Technique for Producing Small Disturbances in Supersonic Wake Flows

J. M. Kendall, Jr.

The processes which initiate turbulence in hypervelocity wake flows are under intensive study by several investigators. The primary process is the aerodynamic amplification of small random flow fluctuations which are naturally present. The instability responsible for the amplification has been studied theoretically but not experimentally for high-speed flows. The purpose of this article is to describe a technique for the production of controlled frequency disturbances in the wake of supersonic wind tunnel models. The rate of downstream growth, which depends on frequency, can be directly measured by means of a hot-wire anemometer.

The method consists of weakly ionizing the flow near the base of the model by maintaining a glow discharge between the metal model and an insulated wire. A fluctuating electrostatic field which drives the ions back and forth is generated by a second wire in close proximity to the first. One successful geometry is shown in Fig. 8.

In this case the model is a two-dimensional plate 0.04 in. thick by 3-in. chord. It spans the 18-in. width of the JPL supersonic wind tunnel and is held under tension. The two wires, 0.01-in. tungsten, are affixed to the model by means of epoxy cement. They terminate within the epoxy near one end of the model, while at the other they extend through the tunnel wall to connection terminals. The tunnel is operated between Mach 2 and 4 with a static pressure of about 0.1 cm Hg. Under these conditions, the glow discharge draws about 3 ma at 500 v from a small regulated power supply. The electrostatic deflection wire is connected to the 600-Ω output of a 200-w audio amplifier whose response is uniform up to about 100 kc. The amplifier is driven by an audio oscillator or white-noise generator.

The amplitude of the resulting disturbance has been estimated to be of the order of 0.05–0.1% mass flow fluctuation. It was found that this amplitude was not large enough to overcome the disturbance caused by sound radiated from turbulent tunnel wall boundary layers. Therefore, the tunnel was operated at the low pressure indicated above in order to obtain laminar wall boundary layers in all cases. Then, introducing disturbances into the wake resulted in a dramatic increase in the wake width when viewed by the schlieren method. It is expected that the technique will be suitable for measuring the stability of axisymmetric wakes as well as boundary layers.

References

1. Mack, L. M., "The Finite Reynolds Number Stability of the Cooled Laminar Boundary Layer at $M_1 = 5.8$," *SPS 37-31*, Vol. IV, pp. 241–244, Jet Propulsion Laboratory, Pasadena, California, February 28, 1965.

2. Christiansen, W. H., "A Seeded-Gas Plasma for Low-Density Wind-Tunnel Use," *Review of Scientific Instruments*, Vol. 36, No. 1, pp. 71–75, January 1965.

3. Stone, P. M., "Cesium Oscillator Strengths," *Physical Review*, Vol. 127, No. 4, pp. 1151–1156, August 15, 1962.

4. Grewal, M. S., and Talbot, L., "Shock Structure in a Partially Ionized Gas," *Journal of Fluid Mechanics*, Vol. 16, Part 4, pp. 573–594, August 1963.

5. Sherman, F. S., "Shock-Wave Structure in Binary Mixtures of Chemically Inert Perfect Gases," *Journal of Fluid Mechanics*, Vol. 8, Part 3, pp. 465–480, July 1960.

XXI. Physics

A. The Superconducting Transition of a Hollow Cylinder Rotating in a Magnetic Field

A. F. Hildebrandt

Earlier reports (Refs. 1 and 2) have described the fluxoid obtained in a hollow tube when it is made superconducting while rotating. The fluxoid Φ is defined by

$$\oint \left[m\,\mathbf{V}_s + \frac{e}{c}\mathbf{A} \right] \cdot d\boldsymbol{\lambda} = \frac{e}{c}\Phi \qquad (1)$$

where the notation is standard, and the contour is any curve in the superconductor that encircles the inner surface. Φ is fixed at the time the cylinder becomes superconducting. The first part of the integral is the mechanical momentum of the super electrons, and the second term includes the flux, which is the magnetic induction times the area. Since the sum is conserved, it follows that the flux (and the induction $\mathbf{B}$) in the bore is a function of the angular velocity of the tube. Only when the tube is at rest is the flux equal to the fluxoid. When the transition occurs with the tube at rest in zero field then Φ is zero.

When the transition occurs while the cylinder is rotating in the presence of a field, then (Refs. 1 and 2) the axial field in the bore should depend on the angular velocity as

$$\mathbf{B} = \mathbf{B}_\infty + \frac{2mc}{e}(\boldsymbol{\omega}_0 - \boldsymbol{\omega}) \qquad (2)$$

where $\mathbf{B}_\infty$ is the applied field, $2mc/e$ the gyromagnetic ratio of the electron, and $\boldsymbol{\omega}_0$ the angular velocity at the time the cylinder became superconducting. This result leads to a minimum in the free energy (Ref. 2). On the other hand, where there is an incomplete Meissner effect this result need not always be observed. For this reason a special case of Eq. (2) will be considered.

If the initially applied field is

$$\mathbf{B}_\infty = -\frac{2mc}{e}\boldsymbol{\omega}_0 \qquad (3)$$

then no fluxoid will be trapped, and the behavior of the tube will be the same as though the transition had occurred with it at rest in zero field. Thus the initially applied field will be observed in the hollow at $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ and zero field at $\boldsymbol{\omega} = 0$.

The consideration of the dynamics of the transition which led to Eq. (2) will be reviewed for the present case. Since the external field is equal to the London field, no surface current will be present on the local sites at which the transition occurs first. This results in zero current on the inner and outer wall of the tube, or a vanishing fluxoid. In other words, there are no shielding currents on the inside or outside of the tube, but a uniform axial field permeating the tube's wall and its bore when the transition is complete.

As the external field is reduced to zero, a shielding current builds up on the outer surface, and the field in the interior is unaffected except for end effects. If the tube is subsequently brought to rest, the shielding current on the outside goes to zero as does the field in the bore. The current on the inner surface (the fluxoid) is always conserved once the transition has occurred. In this case it is zero.

An experiment has been performed to verify this zero current case. The sample consisted of 0.012-cm lead foil formed into a tube 6.4 cm long and 5.0 cm in diameter. The seam was fused by heat. The tube was slipped over a micarta tube and held in place with Ducco clear cement.

The experiment was run in an 8-ft-cubed shielded enclosure, where a field below 0.5×10^{-5} gauss was achieved as described earlier (Refs. 1 and 2). The sample was first cooled to the transition point and then warmed just above the transition in order to assure a minimum cool-down time and associated drift of field. A field was then added in the chosen direction of rotation and of such a magnitude to satisfy Eq. (2). The sample was then rapidly cooled through the transition by adding liquid helium while the speed was kept constant to within a few percent. The flux-gate sensing probe was kept external to the experiment in order to avoid trapping any ac probe fields. After the sample was superconducting, a plot of the axial field was obtained both for $\omega = 0$ and $\omega = \omega_0$. The transition occurred in a field of 7.2×10^{-5} gauss for $\omega_0 = 2\pi \cdot 100$ rps and parallel to the field. The plot is shown in Fig. 1.

The figure shows that the field in the bore, for $\omega = \omega_0$, is very near the applied field and quite uniform. Also, when the sample is brought to rest, the interior field vanishes within experimental accuracy as expected. Other runs were all in qualitative agreement with this result. It was found that if the speed fluctuated even a small amount, the final result could be affected an appreciable amount. This suggests the use of a synchronous (but nonmagnetic) motor. It would also be preferable to run the experiment with tin (transition temperature 3.7°K) in a sealed system where several points could be obtained from a single transfer of liquid helium by alternately pumping and warming the helium. This is not possible with the present apparatus.

When measurements were made with the initial **B** and ω_0 oppositely directed, sharp undirectional axial gradients were observed. Furthermore, it was difficult in this case to obtain reproducible behavior. These results indicate that the value of the fluxoid varied along the length

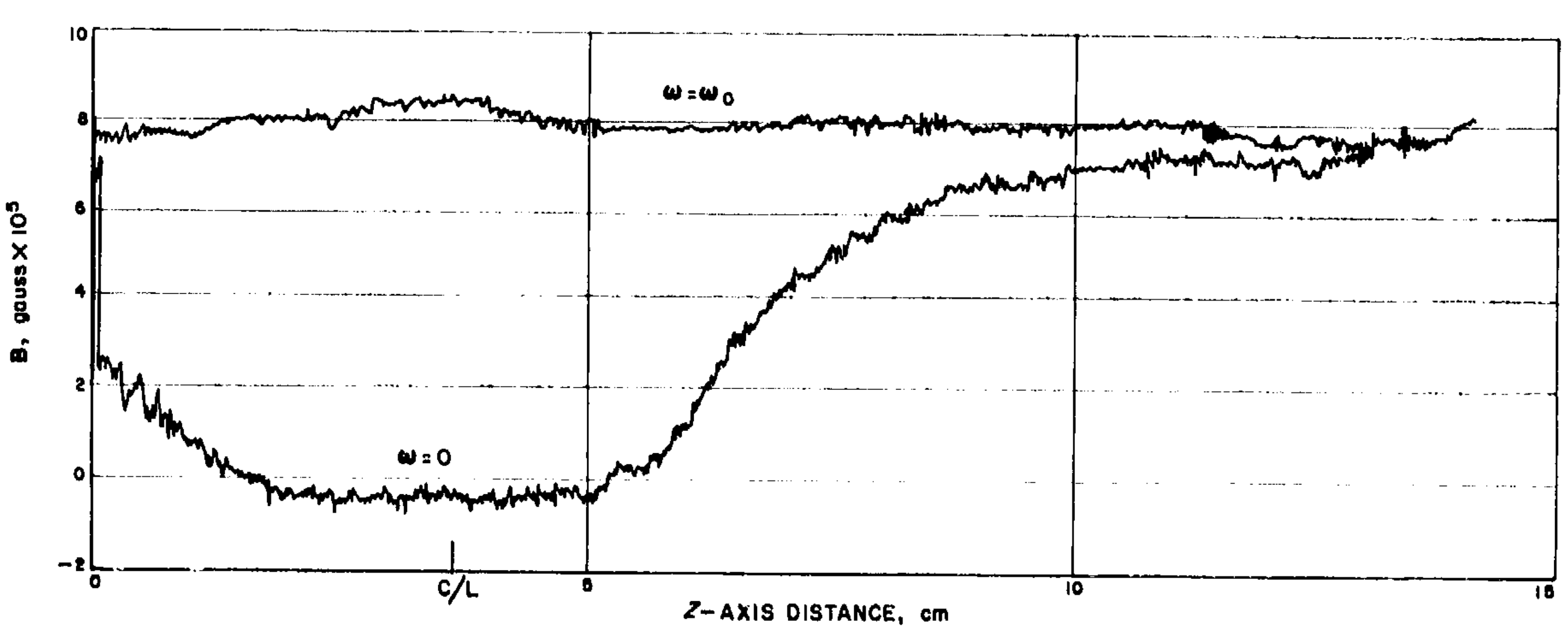


Fig. 1. B–Z plot of the axial field in a hollow lead tube for angular velocity $\omega = 0$ and $\omega = \omega_0$

of the tube. Consideration of this feature indicates that the exact nature of how the transition progresses is important for the generation of such gradients. This suggests studies with pure seamless tin tubes, where the flux leakage through the wall may be kept to a minimum

These results show that zero frozen-in flux should result for a *solid* that is cooled while rotating in the presence of an applied field equal to the London field. In such a solid, the transition occurs first in a thin outer shell and then proceeds inward. For the proper value of applied field this will correspond to zero surface current, and the London value will be the final result. This result is similar to that of observing the Meissner effect for zero field in a non-rotating case; i.e., the uniqueness of the London moment can only be shown if the experiment involves an expulsion of flux or current from the sample as the transition proceeds. For if one considers the transition of a rotating perfectly inviscid electronic fluid (Ref. 3) in the presence of an applied field equal to the London field, then the field in the solid will be the London field (Ref. 4).

If the external field is zero, a non-vanishing fluxoid will be established as the first outer shell goes superconducting, as shown earlier. This fluxoid is then compressed inward as the transition proceeds with a resultant moment of zero in the rotating frame. If the Meissner effect were complete at low fields then the fluxoid would be expelled and the London value would be obtained. It is possible to establish if a sample will exhibit the unique London moment by investigating the Meissner effect in small fields for non-rotating frames. If the Meissner effect is complete in small applied fields, one should be able to observe the unique London moment of a rotating superconductor.

B. Space-Time Symmetry and Mass Splitting

M. M. Saffren

Recently McGlinn (Ref. 5) has noticed that under rather general circumstances a symmetry group that commutes with the homogeneous Lorentz group also commutes with the inhomogeneous Lorentz group (Poincare group) as well. This fact has the important consequence that the mass operator also commutes with the symmetry group so that a set of particles that transform as an irreducible basis of the symmetry group must all have the same mass. This last consequence is not in accord with observation. Accordingly, we must either give up Lorentz invariance, or give up the possibility of having symmetry groups in elementary particle physics other than the Lorentz group; any other symmetry would only be approximate.

We point out that in proving the theorem referred to in Ref. 5 a crucial assumption is made. This assumption in our view is needlessly restrictive. If we drop this restriction, we can easily find groups that commute with the homogeneous Lorentz group but not with the Poincare group—not with space-time translations. Space-time translations as exemplified by the energy-momentum 4-vector of interacting fields furnish many examples, one of which is given below. In such cases the presence of the symmetry group still allows mass splittings.

However, since in these cases the symmetry group fails to commute with the Poincare group, it would appear that these examples are open to another objection: that the symmetry quantum numbers of a particle change in time since the symmetry group fails to commute with the energy. Nevertheless, if the symmetry quantum numbers are determined by the maximum abelian subalgebra of the symmetry group, and if further, this subalgebra commutes with the generators Poincare group—the energy, in particular—this objection is easily met since the quantum numbers are then time-independent. The mass splittings are then also time-independent. (To this set of quantum numbers we can adjoin the eigenvalues of the Casimir operator of the group if this operator also commutes with the Poincare group.)

We now indicate the crucial restriction imposed in Ref. 5. Let X_γ be the generators of the symmetry group, $M_{\mu\nu}$ the generators of the Lorentz group, and P_μ the generators of space-time translations. In Ref. 5 it is assumed that the commutator of any two of these generators is expressible as a linear combination of all others, i.e., the algebra of the generators is a Lie algebra. If we drop this restriction, we are no longer bound by the theorem, as we now show. Let X^0_γ be a member of the maximal abelian subalgebra of the X_γ, and let

$$[X^0_\gamma, P_\mu] = 0$$

$$[X_\gamma, M_{\mu\nu}] = 0$$

In fact, let

$$P_\mu = \sum_\gamma X^0_\gamma (p_\mu)_\gamma$$
$$[X_\delta, (p_\mu)_\gamma] = 0$$

where $(p_\mu)_\gamma$ is a 4-vector operator. Then in general

$$[P_\mu, X_\gamma] \neq 0$$

In fact,

$$[X_\delta, P_\mu] = \sum [X_\delta, X^0_\gamma] (p_\mu)_\gamma$$

and $M^2 = P_\mu P^\mu$ is a function of X^0_ν, i.e., $M^2 = M^2 (X^0_\gamma)$ and we have mass splitting.

As an example, let

$$P_\mu = (p_\mu)_0 + (p_\mu)_I + (p_\mu)_{\tau_3} \tau_3$$

where $(p_\mu)_0$ is the energy-momentum 4-vector of the nucleon and photon fields without interaction between these two fields. Here τ_3 denotes the 3-component of the isotopic spin, and $(p_\mu)_I + (p_\mu)_{\tau_3} \tau_3$ denotes the photon-nucleon interaction. Then

$$[M_{\mu\nu}, \tau_3] = [M_{\mu\nu}, \tau_\pm] = 0$$
$$[P_\mu, \tau_3] = 0$$
$$[P_\mu, \tau_\pm] = a_\pm \tau_\pm (p_\mu) \tau_3$$

where $a_\pm$ = constant, and

$$[M^2, \tau_3] = 0$$
$$[M^2, \tau_\pm] \neq 0$$
$$[M^2, \tau^2] = 0$$

Thus

$$M^2 = M^2 (\tau_3, \tau^2) = M^2_0 + M^2_1 (\tau^2) + \tau_3 M^2_2 (\tau^2)$$

The first term is the mass operator of the non-interacting fields. The second is a mass shift due to the interaction. The last term is a Lorentz invariant mass splitting due to the interaction.

The point here is that the algebra of the nucleon-photon field is not encompassed by the Lie algebra of the Poincare and isotopic spin groups alone. The algebra of the nucleon-photon field is in fact infinite and is comprised of the infinite number of creation and destruction operators of the fields and, of course, their products. Apparently, it is these other parts of the algebra—these other "degrees of freedom"—that allow the mass splitting.

C. Rigid Motions in Einstein Space

H. D. Wahlquist and F. B. Estabrook

In 1908 Max Born (Ref. 6) proposed a Lorentz-covariant kinematic definition of rigidity in special relativity based on constancy of proper distance between adjacent material points. His definition seems appropriate for a rigid body, since it is mathematically equivalent to the vanishing of the rate-of-strain tensor $\sigma_{\mu\nu}$. If λ^μ is the velocity 4-vector of a continuum, the mathematical expression of this condition in covariant notation reads

$$\sigma_{\mu\nu} = \lambda_{\mu,\nu} + \lambda_{\nu,\mu} + a_\mu \lambda_\nu + a_\nu \lambda_\mu = 0 \qquad (1)$$

where a_μ is the acceleration vector, $a_\mu = \lambda_{\mu,\nu} \lambda^\nu$. Born's constraint may also be geometrically characterized by the statement that the metric of the local 3-space (quotient space) orthogonal to the world lines of the rigid body is constant throughout time.

It was soon demonstrated independently by G. Herglotz (Ref. 7) and F. Nöther (Ref. 8) that in the Minkowski space-time of special relativity this constraint leads to the surprising result that a rigid body, so defined, has in general only three degrees of freedom. Practically all allowed rigid motions are non-rotating. The very few special motions with rotation have the property that both the absolute acceleration and angular velocity vectors at each point of the body are constant in time relative to the body. A familiar example is the rigid disk, rotating about a fixed axis, which is compatible with Born's constraint only if the angular acceleration vanishes. This latter class of motions is in fact included in the 10-parameter family of isometries of Minkowski space-time —transformations which carry the entire 4-dimensional manifold into itself. This result, stating that the only allowed rigid motions with rotation are isometries of the space-time manifold, is today often referred to as the Herglotz-Nöther theorem.

Their proof of the theorem for flat space-time led to the conclusion that the ordinary physical concept of rigidity cannot be successfully incorporated in special relativity. Little further work on the subject appeared until recently, when its possible importance for the general theory of relativity was suggested [Synge (Ref. 9)]. The constraint as written in Eq. (1) may be carried over directly into general relativity. Such rigid motions in curved space-time are of considerable interest: mathematically, for their use in the invariant geometric analysis of the structure and global properties of Riemannian manifolds, analogous to that based on the more restricted isometries; and physically, both for the possibility of generating new exact solutions of the Einstein field equations, and for attaining a better understanding of the imprecise concept of rigidity so often implicitly involved in the interpretation of relativistic experiments, e.g., the Pound-Rebka red-shift experiment.

The investigation of possible Born-rigid motions in general relativity divides rather naturally into two different categories. First and simplest is the study of *test* rigid bodies—those which have zero, or at least negligible, stress-energy densities and so do not themselves contribute to the space-time curvature. And second, the study of the *dynamical* rigid body having a non-neglectable stress-energy tensor. A further important distinction must be made according to whether the space occupied by the rigid body contains stress-energy arising from other physical entities, in addition to any associated with the rigid body itself. We shall here adopt the attitude that situations having superimposed foreign stress-energy, through which the rigid body moves, are of lesser physical interest and so exclude them. The problem of *test* rigid bodies is accordingly restricted to empty or Einstein spaces, i.e., those with vanishing Einstein tensor. In the *dynamical* case we are led to require only that the velocity 4-vector of the rigid body is the time-like eigenvector of the Einstein or stress-energy tensor. This condition ensures the vanishing of the local, proper momentum-density vector inside the rigid body and is also, of course, satisfied trivially in the Einstein space, with or without cosmological constant. Thus it will hold generally for both the test and dynamical situations of interest. Since this condition is implicit in most of what follows, it will be assumed here once and for all in order to avoid repetitious qualification of statements.

Some of the questions which may be posed for study are: Does either the test or dynamical rigid body have more than the three degrees of freedom permitted in flat space-time? Is the time behavior of acceleration and angular velocity vectors restricted, and if so, how? Will an arbitrarily given Einstein space admit a test rigid body motion? Will such a space (empty, but perhaps with cosmological constant) admit *rotating* rigid motions other than isometries—that is, does the Herglotz-Nöther theorem still apply? In 1959, C. B. Rayner (Ref. 10) showed that the proper energy density and the *magnitude* of the angular velocity of a rigid body must be constant in time. Recently, Pirani and Williams (Ref. 11) have obtained further results showing that the time constant orthogonal metric of the rigid body leads to a relation between the time behavior of the angular velocity vector of the body and certain components of the space-time curvature. R. H. Boyer[1] has extended this work and used it to obtain *sufficient*, but not necessary, conditions for the validity of the Herglotz-Nöther theorem in curved manifolds.

We have been studying this subject as an application of our dyadic formulation of general relativity (Ref. 12). The 3 + 1 partitioning of space-time relative to a congruence of world lines effected by the dyadic approach permits us to exploit fully the strong geometrical condition on the orthogonal 3-space metric given by Born's constraint. We are thus able to recover all the results of previous authors quickly and transparently, and have developed further equations which are presently being investigated. Several equations from Ref. 12 enter the derivations, but most pertinent is the generalized Gauss equation (Eq. D.34 of Ref. 12) which relates the symmetric curvature tensor (or better, dyadic) E of the quotient 3-space, six of the twenty 4-space curvature components comprised in a symmetric dyadic P, and the angular velocity vector $\boldsymbol{\Omega}$ of the rigid body. The equation, specialized to the case of rigid motion and body-fixed axes, is

$$\mathsf{E} = -(\mathsf{P} + 3\,\boldsymbol{\Omega}\boldsymbol{\Omega}) \tag{2}$$

The constraint of rigidity leads immediately, by other equations of the complete dyadic set, to $\dot{\mathsf{E}} = 0$, expressing time constancy of the orthogonal metric, and to $(Tr\mathsf{P})^{\cdot} = 0$, giving the first of Rayner's results, i.e., constancy of the proper energy density (which is equal to $-1/2\ Tr\mathsf{P}$). The dot denotes the derivative with respect to proper time in body-fixed axes. We have then from Eq. (2),

$$\dot{\mathsf{P}} + 3\dot{\boldsymbol{\Omega}}\,\boldsymbol{\Omega} + 3\boldsymbol{\Omega}\,\dot{\boldsymbol{\Omega}} = 0 \tag{3}$$

which is the relation first obtained by Pirani and Williams. Taking the trace of Eq. (3) now recovers the second of Rayner's results, the constancy of the magnitude of the angular velocity:

$$\boldsymbol{\Omega}\cdot\dot{\boldsymbol{\Omega}} = \frac{1}{2}\,(\Omega^2)^{\cdot} = 0 \tag{4}$$

At this point other dyadic equations are used to develop a lemma whose derivation is elementary but which we shall here simply state: If we are given that $\dot{\boldsymbol{\Omega}} = 0$, but $\boldsymbol{\Omega} \neq 0$, it then follows that the acceleration vector of each point of the body is also independent of time, $\dot{\mathbf{a}} = 0$, and further that $\nabla \times \mathbf{a} = 0$. This lemma is true for both *test* and *dynamical* rigid bodies in *arbitrary* space-time manifolds, subject only to the vanishing of the momentum

[1]Boyer, R. H., *Proceedings of the Royal Society (London)*, (in press); the authors are grateful to Professor Boyer for providing a preprint of this work.

density vector, which has been assumed throughout. However, these properties of the acceleration vector, together with rigidity, define the necessary and sufficient conditions for an isometry of the space-time manifold.

In flat space-time by definition all components of the Riemann tensor, including P, vanish. One has then immediately from Eq. (3), $\boldsymbol{\Omega} = 0$, and, using the lemma, the original Herglotz-Nöther theorem for flat space-time results. The fact that, by means of the lemma, the condition $\dot{\boldsymbol{\Omega}} = 0$ is sufficient to establish the Herglotz-Nöther theorem for an arbitrary manifold is the basis for Boyer's sufficiency conditions, which in our notations are simply $\mathsf{P} - ⅓(Tr\mathsf{P})\mathsf{I} = 0$, leaving the other curvature components arbitrary. The problem of necessary and sufficient conditions for validity of the Herglotz-Nöther theorem is still open. We have, however, now obtained several additional equations relating the space-time curvature to the acceleration and angular velocity vectors of the rigid body, which should bring this problem much closer to solution.

These new equations result from appropriate further space and time differentiation of the relations already presented, in a way that permits introduction of the Bianchi identities to eliminate space derivatives wherever possible. They are expressed in terms of the following quantities: the vectors **a** and $\boldsymbol{\Omega}$ and the 3-space curvature dyadic E, previously introduced; a symmetric, traceless, dyadic B incorporating five of the ten components (the so-called "magnetic" components) of the Weyl tensor; the local, proper stress dyadic T of the rigid body; and finally, the symmetrized curl of T, denoted by G and defined by

$$\mathsf{G} = \nabla \times \mathsf{T} - \mathsf{T} \times \nabla$$

in the usual Gibbs vector-dyadic notation. Adding the new relations, which consist of one scalar, one vector, and one traceless symmetric dyadic equation, to the prior equations relating these quantities, we obtain

$$(\Omega^2)^{\cdot} = 0 \qquad \dot{\mathsf{E}} - (\mathsf{P} + 3\boldsymbol{\Omega}\,\boldsymbol{\Omega})^{\cdot} = 0$$

$$\{\mathsf{B}:\mathsf{B} - 5\mathbf{a}\cdot[\boldsymbol{\Omega}\cdot\mathsf{B} + (\mathbf{a}\cdot\boldsymbol{\Omega})\boldsymbol{\Omega} - 2(\Omega^2)\mathbf{a}]$$
$$- 4\boldsymbol{\Omega}\cdot\mathsf{E}\cdot\boldsymbol{\Omega} + 6\,\boldsymbol{\Omega}\cdot\mathsf{T}\cdot\boldsymbol{\Omega} - (\Omega^2)Tr\mathsf{T}\}^{\cdot} = 0$$

$$\{\boldsymbol{\Omega}\cdot\mathsf{B} + (\mathbf{a}\cdot\boldsymbol{\Omega})\,\boldsymbol{\Omega} - 2(\Omega^2)\mathbf{a}\}^{\cdot} = 0$$

$$\{\dot{\mathsf{B}} - \boldsymbol{\Omega}\times\mathsf{B} + \mathsf{B}\times\boldsymbol{\Omega} + \mathbf{a}\times\mathsf{E} - \mathsf{E}\times\mathbf{a}$$
$$+ 3\boldsymbol{\Omega}\,\boldsymbol{\Omega}\times\mathbf{a} - 3\mathbf{a}\times\boldsymbol{\Omega}\,\boldsymbol{\Omega}$$
$$- \mathbf{a}\times\mathsf{T} + \mathsf{T}\times\mathbf{a} - \mathsf{G}\}^{\cdot} = 0 \qquad (6)$$

This set of equations applies to both *dynamical* and *test* rigid bodies. For the case of test bodies in Einstein space, important simplification takes place. Firstly, the dyadics T and G must both vanish. Secondly, P reduces to a traceless, symmetric dyadic A, containing the other five components (the so-called "electric" components) of the Weyl tensor. In this case, then, these equations constitute a set of *fifteen* independent ordinary differential equations relating the time development of the *sixteen* scalar quantities in **a**, $\boldsymbol{\Omega}$, A, and B. Determination of the consequences of these equations for the several types of algebraically special Einstein spaces with cosmological constant is proceeding. Among the results is that, for space-times of constant curvature and of Petrov type *N*, the Herglotz-Nöther theorem holds. Further, in the case of type *N* Einstein spaces, the *only* Born-rigid motions allowed are isometries. A paper detailing the derivation of the equations and examining some of the consequences thereof is in preparation.

References

1. Hildebrandt, A. F., and Saffren, M. M., "The Superconducting Transition of a Rotating Superconductor—The Hollow Cylinder," *Proceedings of the 9th International Conference on Low Temperature Physics*, September 1964; *SPS 37-30*, Vol. IV, December 31, 1964.

2. Saffren, M. M., *The Rotating Superconductor—Part I: The Fluxoid*, Technical Report 32-650 (Part I), Jet Propulsion Laboratory, Pasadena, California, March 15, 1965.

3. Becker, R., Sauter, F., and Heller, C., *Zeitschrift für Physik*, Vol. 85, p. 772, 1933.

References (Cont'd)

4. Saffren, M. M., "The Rotating Superconductor—Part III: The Superelectrons as an Incompressible Inviscid Charged Fluid (BSH) Theory," Technical Report 32-650 (Part III), Jet Propulsion Laboratory, Pasadena, California, (to be published).

5. McGlinn, W. D., *Physical Review Letters*, Vol. 12, p. 467, 1964.

6. Born, M., *Annalen der Physik*, Vol. 30, p. 1, 1909.

7. Herglotz, G., *Annalen der Physik*, Vol. 31, p. 393, 1910.

8. Nöther, F., *Annalen der Physik*, Vol. 31, pp. 919–944, 1910.

9. Synge, J. L., *Relativity: The General Theory*, North-Holland Publishing Co., Amsterdam, 1960.

10. Rayner, C. B., *Comptes Rendus*, Vol. 248, p. 929, 1959a.

11. Pirani, F. A. E., and Williams, G., *Séminaire JANET*, No. 8–9, 5ième année, 1961/2.

12. Estabrook, F., and Wahlquist, H., *Journal of Mathematical Physics*, Vol. 5, p. 1629, 1964.

XXII. Applied Science

A. Infrared Detector Analysis Facility

F. C. Haley and N. F. Stahlberg

1. Introduction

Early in 1963 it was decided that a facility was necessary for the evaluation and calibration of infrared detectors and detection systems. Several detector analysis systems are, and were then, commercially available; however, most are intended for production-line use and are not well suited for the particular, and relatively unusual detector configurations required for spacecraft instruments. For this reason, the decision was made that the analysis system would be built at JPL using selected individual components and employing such special physical configurations as were appropriate to the particular projects at hand. The initial project for which the system was to be used was the *Mariner* 1966 multidetector infrared spectrometer. The system was made to be as versatile and comprehensive as possible so that it might also serve as a general laboratory tool for infrared measurements.

This article describes the resulting facility in its present state of development.

2. Detector Evaluation

a. Criteria of performance. The overall performance of an infrared detector is described by some figure of merit (Refs. 1–3). Figures describing the performance of thermal detectors are generally referred to radiation from a blackbody source, whereas those relating to photon detectors are most meaningful when referred to monochromatic radiation. These latter, spectral figures of merit are those with which we are presently concerned.

Figures of merit presently in general use include the following:

(1) *Responsivity:* The rms signal voltage per unit rms radiant power incident upon the detector.

(2) *Noise equivalent power* (*NEP*): The rms value of the radiant power incident upon the detector which results in an rms signal voltage equal to the intrinsic rms noise voltage of the detector.

(3) *Detectivity:* The rms signal-to-noise ratio per unit radiant power incident upon the detector (reciprocal of NEP).

(4) *Specific detectivity* (D^*): The detectivity normalized to unit sensitive area.

Figures of merit must be accompanied by reference to the source of radiation, modulation frequency, and electrical bandwidth. Very often data are presented as families of curves (e.g., D^* vs λ) at different detector temperatures and bias conditions.

b. Measurable parameters. As has been shown, the calculation of a spectral figure of merit describing detector performance requires the measurement of signal voltage as a function of incident radiant power, wavelength, modulation frequency, bias current, and temperature, and of noise voltage as a function of frequency, bias, and temperature. Preliminary to these measurements, it is necessary to determine the detector impedance as a function of temperature and incident radiant power.

To prepare and evaluate a detector for use at a specific wavelength under given environmental conditions, the overall performance is relatively unimportant; rather, it is necessary to maximize the detector response at the particular wavelength by determining the optimum bias current, load resistance, modulation frequency, and operating temperature.

3. System Description

a. Optical. The source of monochromatic radiation is a Leiss double prism monochrometer, which presently has a spectral range of 0.2 to 9 μ. This range may be extended in either "direction" by use of interchangeable prisms. The monochrometer has an effective focal ratio of about f/7 and exceptionally high spectral purity.

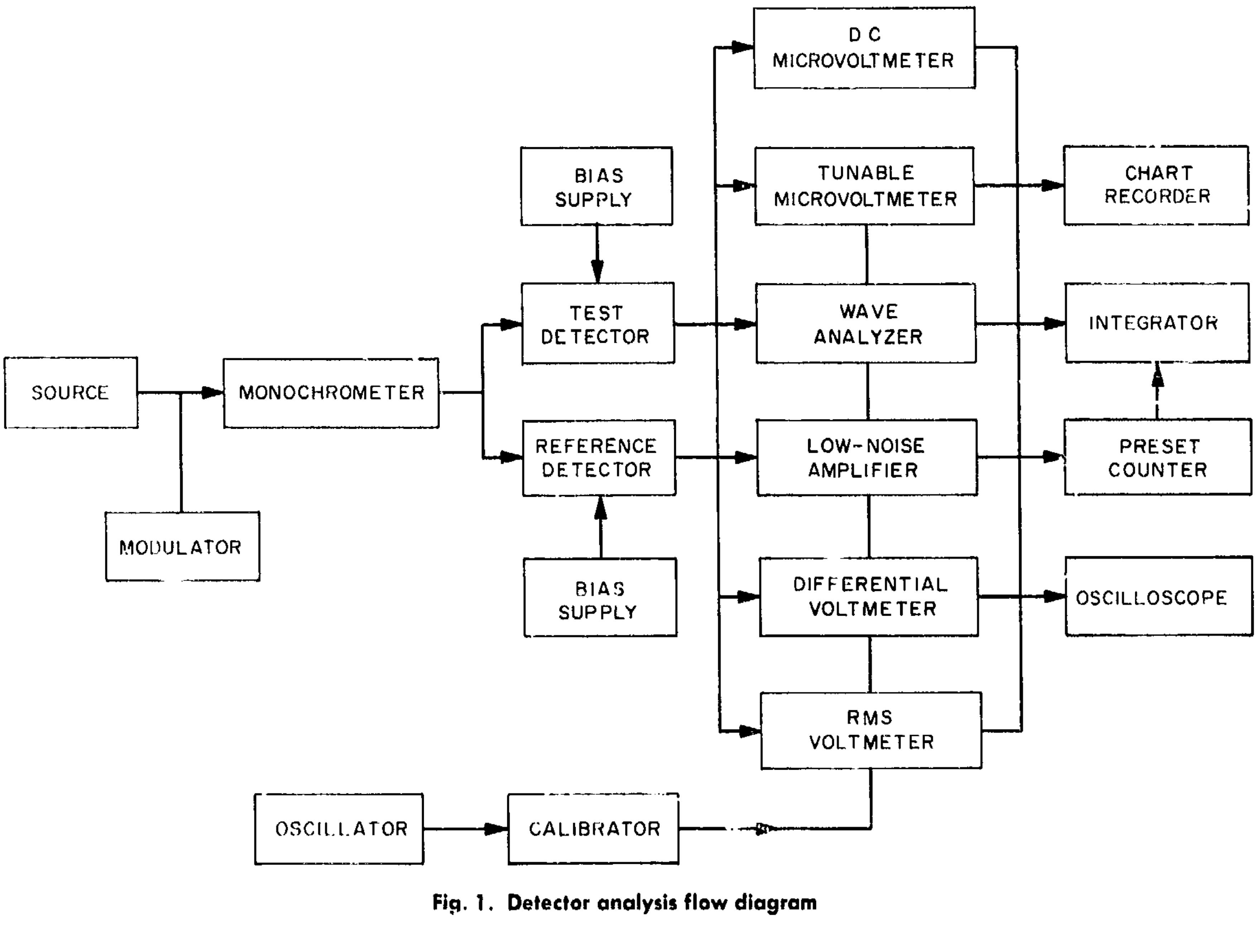


Fig. 1. Detector analysis flow diagram

The instrument is mounted on a traversing platform, enabling it to be positioned alternately before the detector under test and a standard of reference, the latter being either a calibrated thermopile or a Golay detector. The signals from the detectors are fed into amplifiers appropriate to their modes of operation and impedances.

Radiant power is initially provided by a Nernst glower, while means of operating gas-discharge tubes and hollow-cathode lamps are furnished for the purpose of calibration at short wavelengths. Absorption phenomena are employed for long-wavelength calibration.

b. Electronic. Six different signal amplifiers are presently included in the system, their respective capabilities being as follows:

(1) Dc microvoltmeter with 0.1-μv to 1-v measurement capability; input impedance of 1 MΩ and greater, depending on range.

(2) Tunable microvoltmeter (non-synchronous) with 0.1-μv to 15-mv sensitivity, 1-cps to 1-kc frequency range with variable passband; input impedance selectable from 10 Ω to 40 MΩ.

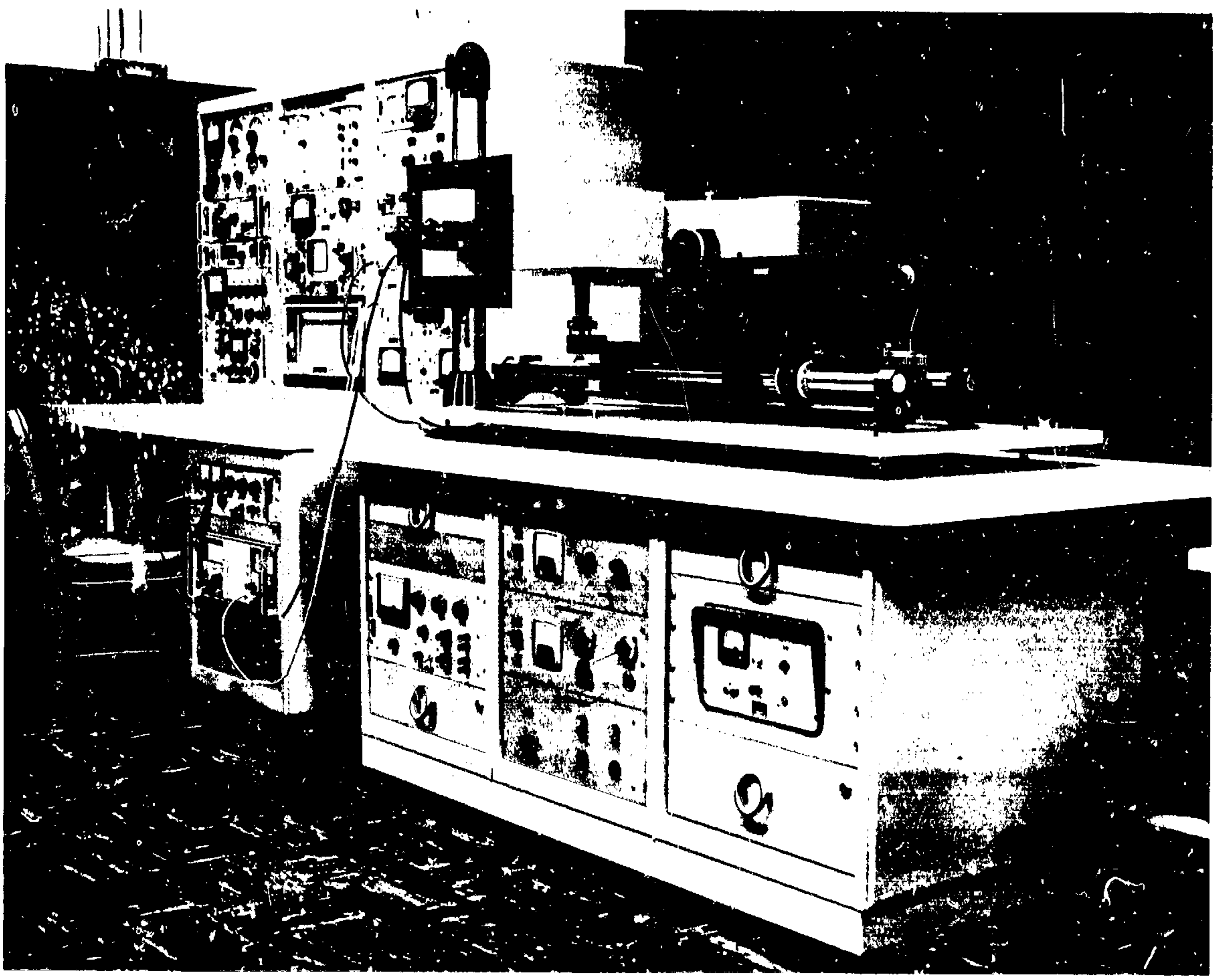


Fig. 2. Infrared detector analysis system

(3) Wave analyzer with 10-μv to 300-v sensitivity, 20-cps to 50-kc frequency range, fixed passband; 100-kΩ input impedance.

(4) True rms voltmeter with 0.1-mv to 300-v sensitivity, 10-cps to 10-Mc frequency range; 10 MΩ input impedance.

(5) Differential voltmeter with ac sensitivity of 1 mv to 500 v and input impedance of 1 MΩ.

(6) Low-noise amplifier with maximum gain of 1000, 10-cps to 10-kc frequency range, and selectable high- and low-pass filters.

The true rms and differential voltmeters may be preceded by a bandpass filter having a frequency range of 0.02 cps to 2 kc. The differential voltmeter has the capability of suppressing a constant noise component which appears to be non-filterable. This instrument is also used as a calibration standard for dc voltage. Final data may be presented on a strip-chart recorder or through an integrator having selectable integration times of from 2 to 120 sec.

Four dc power supplies are provided for detector bias and for powering any associated electronic circuitry.

Also included in the system are an oscillator and voltage divider for ac voltage calibration; a preset counter for frequency monitoring and precise triggering of integration circuitry; an oscilloscope for waveform monitoring; an impedance bridge; and a line voltage regulator for powering all critical instruments.

The operation of the analysis system is shown diagrammatically in Fig. 1. Fig. 2 shows the system in use, comparing the spectral response of a thermopile and a Golay detector. Close-up views of the instrument console and the monochrometer assembly are shown in Figs. 3 and 4, respectively.

4. Conclusion

During the course of development of the system, the desirability of making several changes and additions has become apparent. Amplifiers with higher sensitivity and lower inherent noise would permit more satisfactory performance at low radiant energy levels and allow the system to be truly detector-noise limited. Presently contemplated improvements include a highly sensitive, low-noise synchronous amplifier for ac signal detection; an extremely low-noise dc nanovoltmeter; and a microammeter with zero suppression to allow eventual use with photomultiplier and phototubes. Optical and mechanical modifications are to be made as required by specific projects.

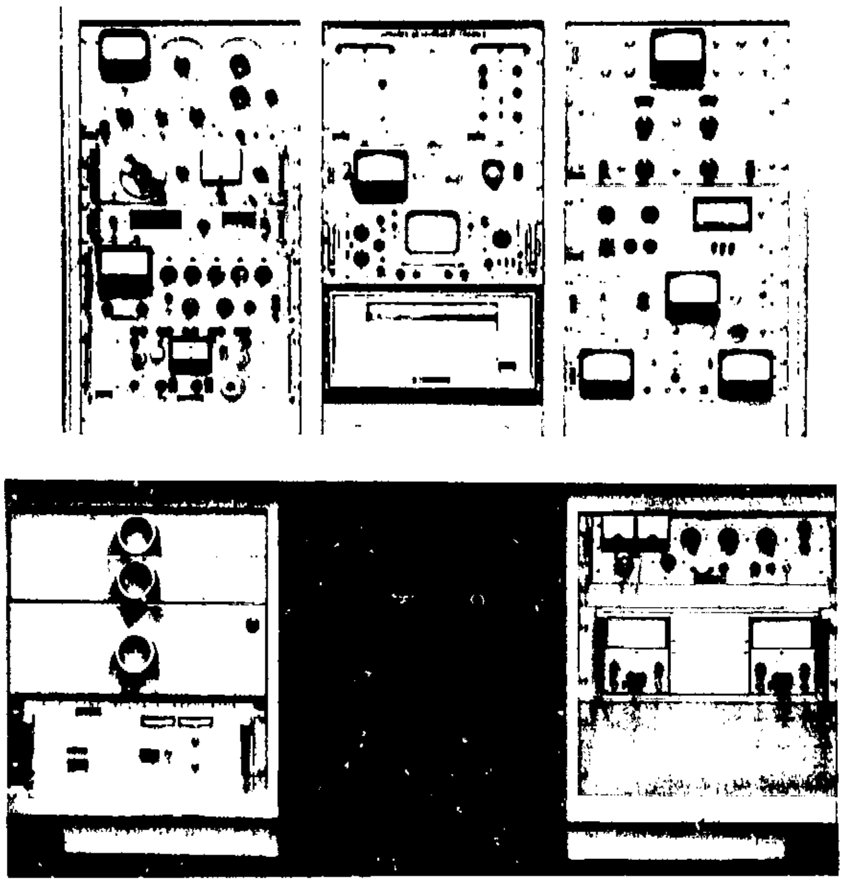


Fig. 3. Instrument console

References

1. Conn and Avery, *Infrared Methods*, Chapter III, Academic Press, New York and London, 1960.
2. Jamieson, Plass, McFee, Grube, and Richards, *Infrared Physics and Engineering*, Chapter 5, McGraw-Hill Book Company, New York, 1963.
3. Kruse, McGlauchlin, and McQuistan, *Elements of Infrared Technology*, Chapters 8 and 10, John Wiley & Sons, New York, 1962.

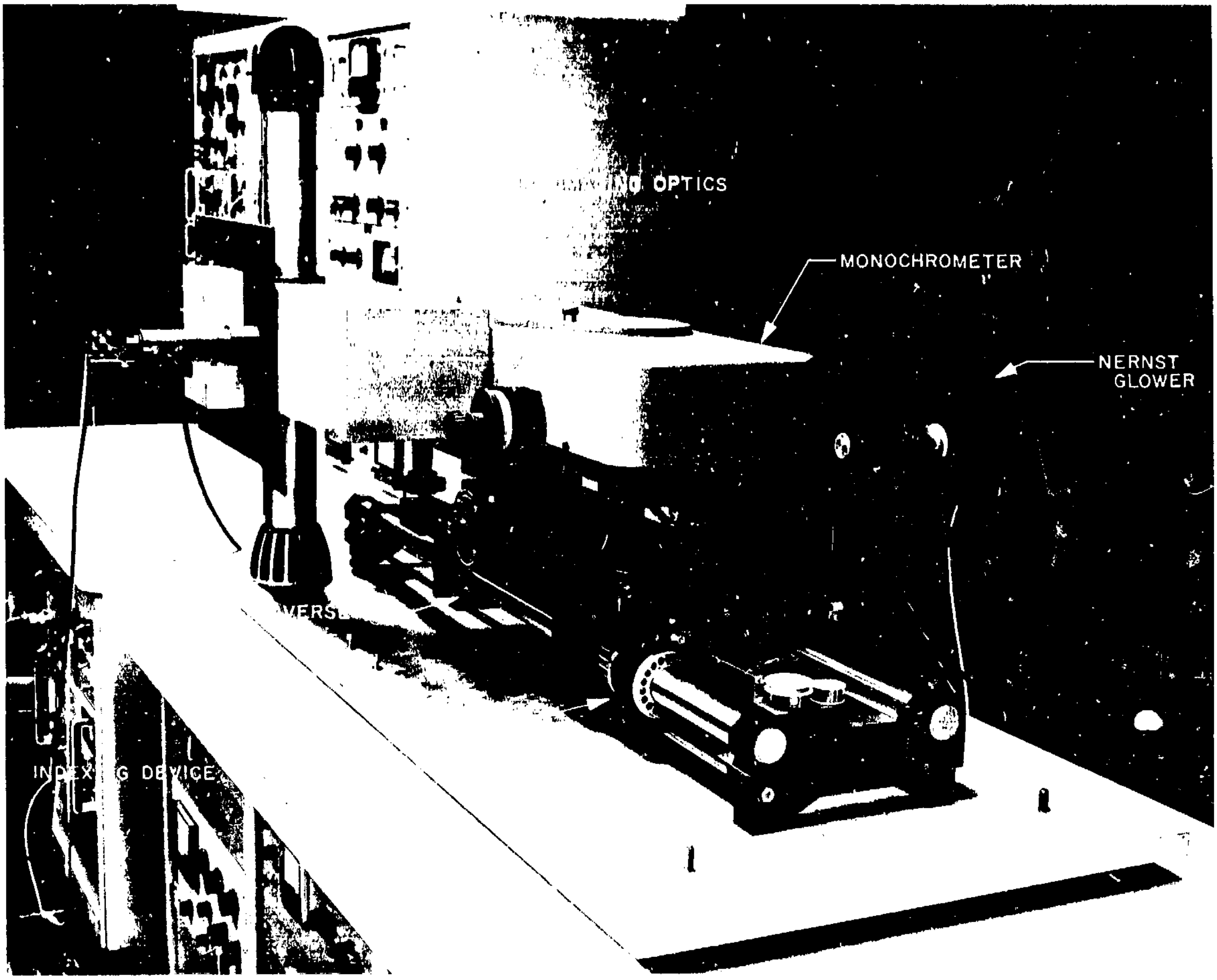


Fig. 4. Monochrometer assembly

BLANK PAGE

TELECOMMUNICATIONS DIVISION

XXIII. Communications Elements Research

A. Low-Noise Amplifiers

W. V. T. Rusch and C. T. Stelzried

1. 90-Gc Millimeter Wave Work,

W. V. T. Rusch and C. T. Stelzried

The objective of the millimeter wave work is to investigate millimeter wave components and techniques to ascertain the future applicability of the frequency range to space communications and tracking. This involves the development of instrumentation for accurate determination of insertion loss, VSWR, power and equivalent noise temperature of passive elements, and gain and bandwidth of active elements at millimeter wavelength. Observations of the Moon taken during the total lunar eclipse on the night of December 18–19, 1964 have been reduced and the data presented.

In order to bring together state-of-the-art millimeter wave circuit elements and evaluate their use in a system, a radio telescope consisting of a 60-in. antenna and a superheterodyne radiometer have been built (*SPS 37-28*, Vol. IV, p. 147 and *SPS 37-29*, Vol. IV, p. 191). The radio telescope was used to observe the 90-Gc temperature of the Moon during the December 30, 1963 eclipse (*SPS 37-26*, Vol. IV, p. 181) and the most recent eclipse, December 18, 1964. These experiments were joint efforts by personnel from JPL and the Electrical Engineering Department, University of Southern California.

Because of unfavorable weather conditions in Pasadena during the latter eclipse, the 90-Gc transportable radio telescope was moved to Palm Springs. Thirteen observations were taken before the end of the umbral phase of the eclipse, and a total of 71 was taken during the night. Both lunar and solar data were taken over a 3-day period for calibration. Data reduction is currently under way to attempt to determine the change in the 90-Gc lunar surface temperature, due to the eclipse.

a. Lunar eclipse results. The 71 observations have been reduced by joint experimenters at the Electrical Engineering Department, University of Southern California. Mid-eclipse occurred at 6.37 p.m., PST; the end of totality

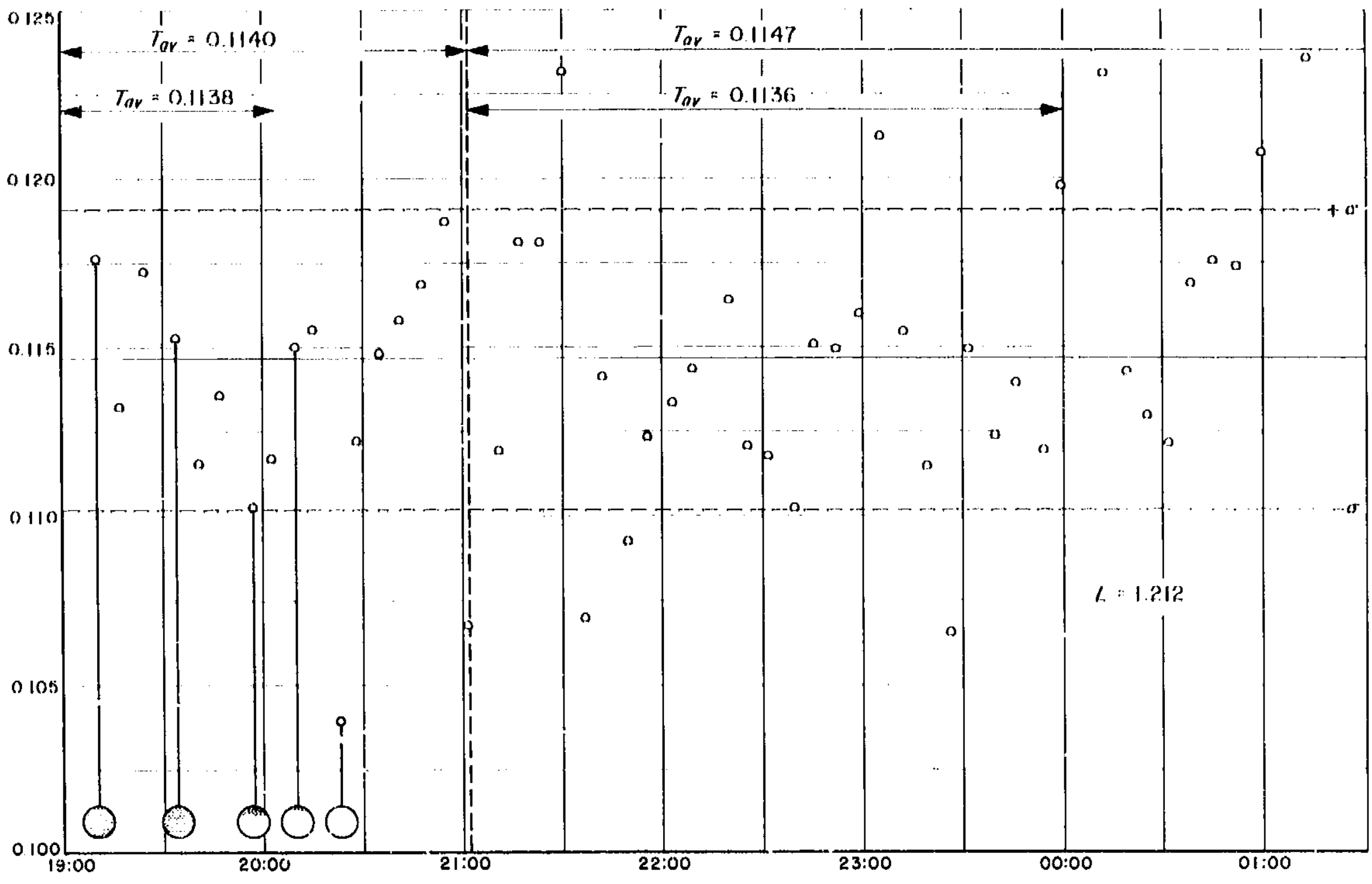


Fig. 1. Equivalent lunar temperature during the night of December 18–19, 1964

was at 7:07 p.m., the end of the umbral phase at 8:15 p.m., and the end of the penumbral phase at 9:14 p.m. The first observation (semidrift curve) was made at the end of totality. Extrapolation of all the data obtained after 9:02 p.m. yielded an atmospheric loss, L of 1.212 at zenith and a Moon temperature of 0.11468 arbitrary units (the arbitrary unit being proportional to the equivalent excess noise temperature of the radiometer gas tube). By multiplying each data point by the factor $(1.212)^{\sec z}$, where z is the lunar zenith angle for that point, it is possible to obtain an equivalent lunar temperature for each data point, commencing with the end of totality and continuing through the night. These points are plotted in Fig. 1 as a function of local time. Limits of $\pm 1\ \sigma$ are indicated. This value is approximately 4%. Relative phases of the eclipse are indicated on the figure. The average value of the data points from 9:02 p.m. until 3:54 a.m. the next morning was 0.1147. The average from 9:02 p.m. until midnight was 0.1136. The average value from 7:11 until 8:03 p.m. was 0.1138 and from 7:11 until 9:02 p.m. was 0.1140.

b. Waveguide calibrations. A block diagram of the millimeter radiometer is shown in Fig. 2. The "hot" loads are temperature-controlled with mercury thermostats. The equivalent noise temperature of the gas tube noise source is being evaluated by comparing against the "hot" loads. The insertion loss of the waveguide components is being measured for this calibration.

2. Continuous Wave (CW) Power Calibration with Thermal Noise Standards, C. T. Stelzried

CW signal power levels can be calibrated using thermal noise standards. This principle can be used to calibrate power levels from the test transmitter used for AGC curves in a low-noise receiver directly at the receiver front end. The technique uses the power ratio measurement method normally used in equivalent noise temperature calibrations.

a. Measurement technique. The calibration technique consists of comparing the total power contained in the

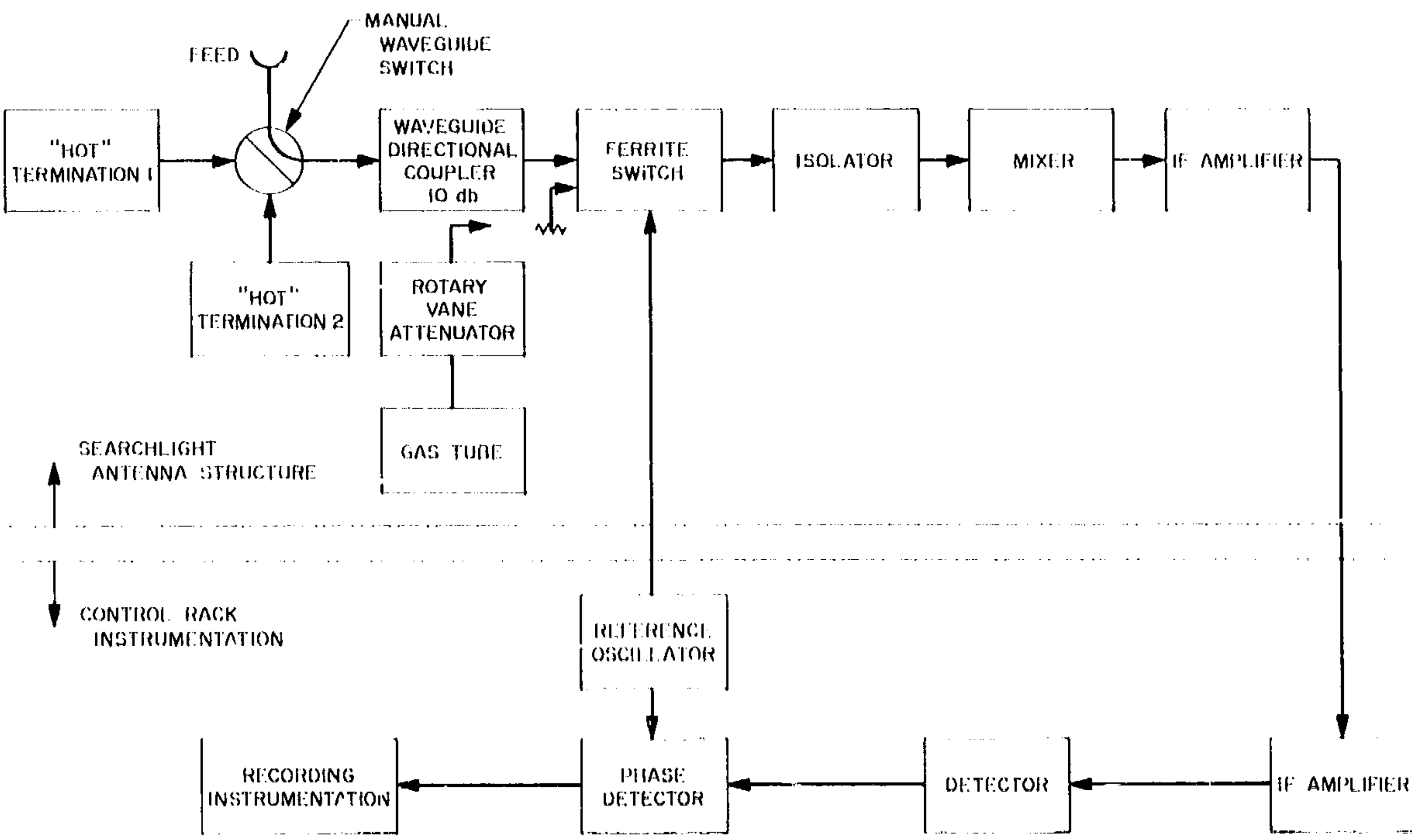


Fig. 2. Millimeter radiometer

system noise P_n plus the CW signal power P_s to the system noise, on the output of the receiving system (Fig. 3). The precision IF attenuator is adjusted for equal power levels with the CW signal on and off. The power ratio is

$$Y = \frac{P_n + P_s}{P_n}. \quad (1)$$

Solving for the signal power,

$$P_s = (Y - 1) P_n. \quad (2)$$

The noise power P_n observed on the filter output with a system temperature T_s and over-all receiver gain G is a function of the filter gain response $G(f)$,

$$P_n = kT_sG \int_0^\infty G(f)\,df. \quad (3)$$

The signal power P_s observed on the filter output is a function of the input signal power P_{si} and the filter gain $G(f_s)$ at the signal frequency f_s and the over-all gain G,

$$P_s = P_{si}G(f_s)\,G. \quad (4)$$

Substituting Eqs. (3) and (4) into (2),

$$P_{si} = (Y - 1)\,kT_s \frac{\int_0^\infty G(f)\,df}{G(f_s)}. \quad (5)$$

Normalizing the filter gain,

$$g(f_s) = \frac{G(f_s)}{g(f_0)}. \quad (6)$$

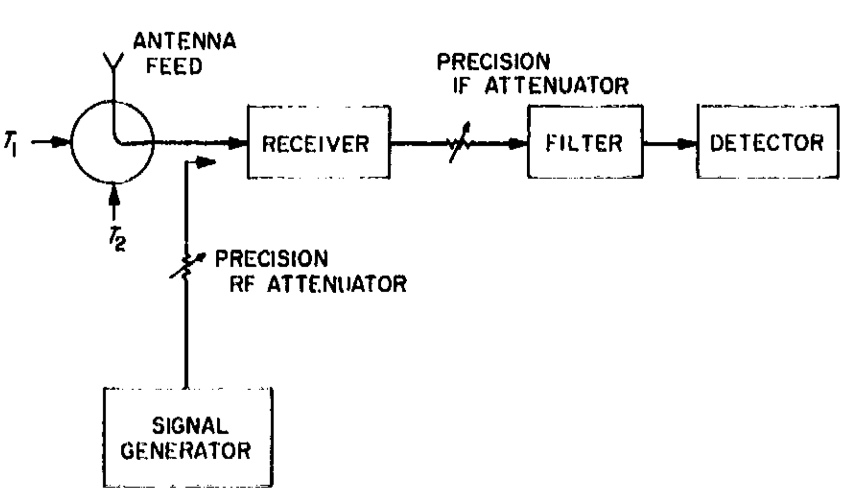


Fig. 3. Receiving system used in evaluation of CW signal

Then

$$P_{si} = (Y-1)\frac{kT_s}{g(f_s)}\,\frac{\int_0^\infty G(f)\,df}{g(f_0)}, \tag{7}$$

The bandwidth is defined (Ref. 1)

$$B = \frac{\int_0^\infty G(f)\,df}{g(f_0)}. \tag{8}$$

Usually f_0 is the frequency at which the filter has a maximum response so that $g(f_0)$ is a maximum. Then if f_s is set equal to f_0, $g(f_s)$ equals 1. Substituting Eq. (8) into (7),

$$P_{si} = (Y-1)\,kT_s\,\frac{B}{g(f_s)}. \tag{9}$$

b. Error analysis. The measurement error involved with Eq. (9) can be determined with (Ref. 2)

$$\left(\frac{\sigma_{P_{si}}}{P_{si}}\right)^2 = \left(\frac{\partial P_{si}/\partial y}{P_{si}}\right)^2 \sigma_y^2 + \left(\frac{\partial P_{si}/\partial T_s}{P_{si}}\right)^2 \sigma_{T_s}^2 + \left(\frac{\partial P_{si}/\partial B}{\partial P_{si}}\right)^2 \sigma_B^2 + \left(\frac{\partial P_{si}/g(f_s)}{\partial P_{si}}\right)^2 \sigma_g^2(f_s) \tag{10}$$

where

σ_Y = standard deviation of the power ratio measurement.

σ_{T_s} = standard deviation of the system temperature measurement.

σ_B = standard deviation of the bandwidth measurement.

$\sigma_{g(f_s)}$ = standard deviation of the filter signal gain measurement.

Substituting the differentials of Eq. (9)

$$\left(\frac{\sigma_{P_{si}}}{P_{si}}\right)^2 = \left(\frac{\sigma_Y}{Y}\right)^2\left(1+\frac{kT_sB}{P_{si}g(f_s)}\right)^2 + \left(\frac{\sigma_{T_s}}{T_s}\right)^2 + \left(\frac{\sigma_B}{B}\right) + \left(\frac{\sigma_{g(f_s)}}{g(f_s)}\right)^2. \tag{11}$$

The standard deviation of the Y factor error is composed of the resolution of the IF precision attenuator σ_a, the input signal level instability $\Delta P_{si}/P_{si}$, and a radiometer instability (Ref. 3) given by

$$[1/\tau B + (\Delta G/G)^2]^{1/2} \tag{12}$$

where

τ = post detector time constant.

Accounting for these,

$$\left(\frac{\sigma_{P_{si}}}{P_{si}}\right)^2 = \left[\left(\frac{\sigma_a}{a}\right)^2 + \left(\frac{1}{\tau B}\right) + \left(\frac{\Delta P_{si}}{P_{si}}\right)^2 + \left(\frac{\Delta G}{G}\right)^2\right] \times \left(1+\frac{kT_sB}{P_{si}g(f_s)}\right)^2 + \left(\frac{\sigma_{T_s}}{T_s}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2 + \left(\frac{\sigma_{g(f_s)}}{g(f_s)}\right)^2. \tag{13}$$

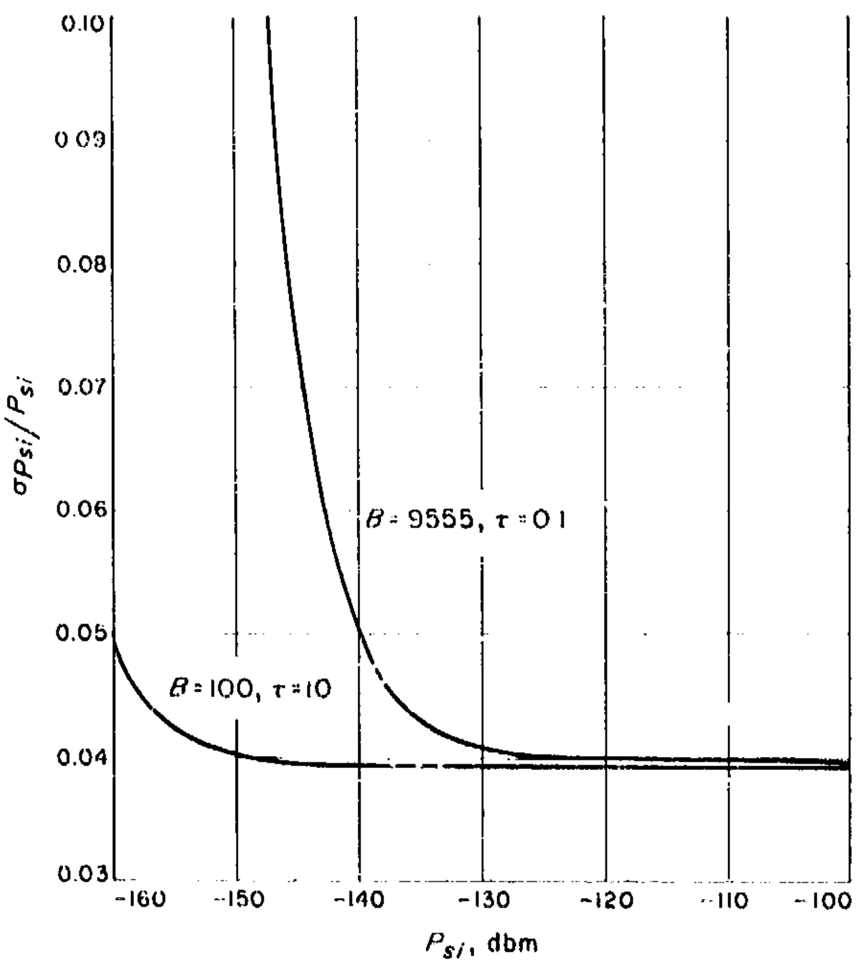


Fig. 4. CW signal power measurement resolution versus signal level

Eq. (13) is plotted in Fig. 4 and two values of τ and B using the values

$$k = 1.38 \times 10^{-23}$$

$$\frac{\sigma_a}{a} = 0.0046$$

$$\left(\frac{\Delta P_{si}}{P_{si}}\right)^2 + \left(\frac{\Delta G_i}{G_i}\right)^2 = 0.0023$$

$$\frac{\sigma_{T_s}}{T_s} = 0.02$$

$$\frac{\sigma_B}{B} = 0.01$$

$$T_s = 27.7^\circ\mathrm{K}$$

$$\frac{\sigma_{g(f_s)}}{g(f_s)} = 0$$

$$g(f_s) = 1$$

Narrowing the bandwidth and increasing the post detector time constant increases the resolution with lower power levels.

3. Equivalent Noise Temperature of Microwave Terminations, C. T. Stelzried

The equivalent input noise temperature derivation for a microwave calibrated termination is reviewed. The exact calibration formula is given. As a practical example, a convenient approximation accounting for first order effects of transmission lines losses, external temperature, frequency deviations assuming a transmission line operated in the TE_{10} mode, and cryogenic temperature pressure dependence is shown to be

$$T' \simeq T(760) + a\Delta P + 0.23026\,[T_0 - T(760)]\left[\frac{L_1}{2}(\mathrm{db}) + L_2(\mathrm{db})\right] \cdot \left(1 + \frac{\Delta f}{f_0}\left\{\frac{\left(\frac{f_0}{f_c}\right)^4 - 6\left(\frac{f_0}{f_c}\right)^2 + 1}{2\left[\left(\frac{f_0}{f_c}\right)^4 - 1\right]}\right\}\right)$$

where the terms are described in the text.

a. Introduction. Microwave terminations are used to calibrate noise sources, receiver temperatures and antenna temperatures. These calibrations are used with radar-radio astronomy observations to determine system sensitivity and calibrations. Calibration precision and long-term stability are prime requirements for the terminations and associated transmission lines.

The equivalent input noise temperature T' for a transmission line of length l terminated with a termination at a temperature T, with an arbitrary temperature distribution T_L and uniform propagation loss per unit length α taken to be independent of temperature is (Ref. 5)

$$T' = Te^{-2\alpha l} + 2\alpha \int_0^l e^{2\alpha(x-l)} T_L(x)\,dx. \tag{1}$$

The transmission line insertion loss is related to α by Ref. 4

$$L = e^{2\alpha l} \tag{2}$$

and

$$2\alpha l = \frac{L(\mathrm{db})}{10\log_{10}e} \simeq 0.23026\,L(\mathrm{db}). \tag{3}$$

Substituting the appropriate temperature distribution along the transmission line, Eq. (1) can be integrated directly. Assuming a linear temperature distribution from T to T_0 and loss L_1 on the first section of the transmission line and a constant temperature distribution T_0 and loss L_2 in the second section of the transmission line

$$T' = T + (T_0 - T)\left[1 - \frac{(L_1 - 1)\,10\log_{10}e}{L_1 L_2 L_1(\mathrm{db})}\right]. \tag{4}$$

With small losses this can be approximated to good accuracy by

$$T' \simeq T + 0.23026\,(T_0 - T)\left[\frac{L_1}{2}(\mathrm{db}) + L_2(\mathrm{db})\right]. \tag{5}$$

If the transmission line insertion loss measurements were made at a frequency f_0, a correction for a new operating frequency $f = f_0 + \Delta f$ can be obtained from

$$\frac{\partial L(\mathrm{db})/\partial f}{L(\mathrm{db})}$$

evaluated at f_0.

Multiplying by Δf gives the first order correction term as in a Taylor's expansion (Ref. 6) for the insertion loss change with small frequency differences. With this addition,

$$T' \simeq T + 0.23026\,(T_0 - T)\left[\frac{L_1}{2}\,(\mathrm{db}) + L_2\,(\mathrm{db})\right]$$

$$\cdot\left[1 + \Delta f \cdot \frac{\partial L\,(\mathrm{db})/\partial f}{L\,(\mathrm{db})}\bigg|_{f=f_0}\right]. \qquad (6)$$

For a rectangular waveguide operated in the TE_{10} mode with a height-to-width ratio of 1:2, the approximate waveguide wall losses are (Ref. 7)

$$L\,(\mathrm{db}) \sim (f)^{1/2}\,\frac{1 + (f_c/f)^2}{[1 - (f_c/f)^2]^{1/2}} \qquad (7)$$

where

$f_c = c/2a =$ cutoff frequency

$c =$ propagation velocity $\simeq 3 \times 10^{10}$ cm/sec

$a =$ waveguide width

Differentiating and dividing by L(db)

$$\frac{\partial L\,(\mathrm{db})/\partial f}{L\,(\mathrm{db})}\bigg|_{f=f_0} = \frac{1}{f_0}\left\{\frac{\left(\dfrac{f_0}{f_c}\right)^4 - 6\left(\dfrac{f_0}{f_c}\right)^2 + 1}{2\left[\left(\dfrac{f_0}{f_c}\right)^4 - 1\right]}\right\}. \qquad (8)$$

The multiplier for $1/f_0$ or frequency deviation sensitivity in Eq. (8) is shown plotted in Fig. 5. Insertion loss changes with frequency are least sensitive at the minimum loss frequency where the derivative is zero ($f_0 \simeq 2.42\,f_c$). This is above the normal operating frequency for most waveguides. Substituting Eq. (8) into (6),

$$T' \simeq T + 0.23026\,(T_0 - T)\left[\frac{L_1}{2}\,(\mathrm{db}) + L_2\,(\mathrm{db})\right]$$

$$\cdot\left(1 + \frac{\Delta f}{f_0}\left\{\frac{\left(\dfrac{f_0}{f_c}\right)^4 - 6\left(\dfrac{f_0}{f_c}\right)^2 + 1}{2\left[\left(\dfrac{f_0}{f_c}\right)^4 - 1\right]}\right\}\right) \qquad (9)$$

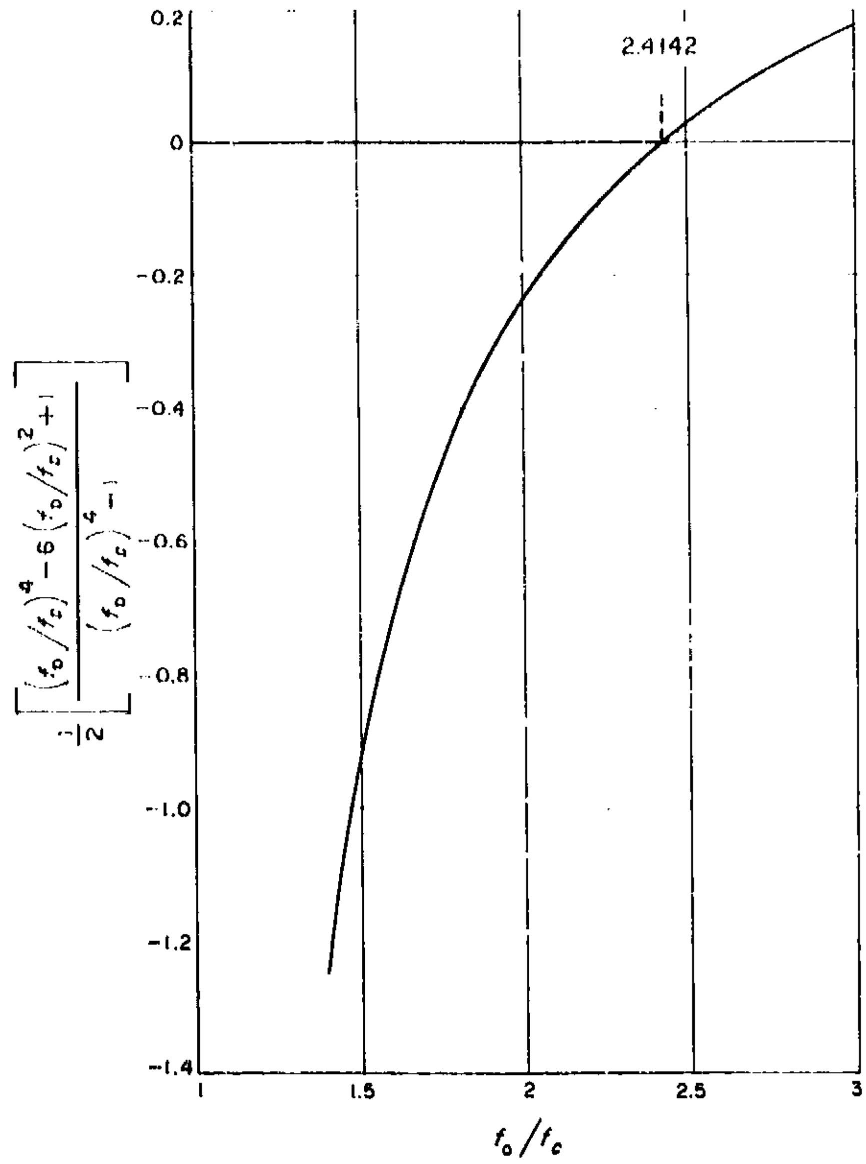


Fig. 5. Plot of insertion loss frequency deviation sensitivity versus frequency

This is equivalent input noise temperature of a calibrated termination with a waveguide transmission line operating in the TE_{10} mode, accounting for ambient temperature, transmission line insertion loss and small frequency deviations. Corrections for frequency deviations are only valid if the VSWR is compatible with the accuracy desired.

If the termination thermal temperature T is determined by a cryogenic liquid (*SPS 37-31*, Vol. III, p. 60), atmospheric pressure must be taken into account to obtain greatest precision. Fig. 6 shows a graph of the boiling temperature of liquid nitrogen as a function of pressure (Ref. 8). This can be approximated by

$$T \simeq T\,(760) + a\Delta P \qquad (10)$$

where $T(760)$ is the liquid temperature at a pressure of 760 mm Hg and ΔP is the pressure deviation in millimeters of mercury. For liquid nitrogen, $T(760) = 77.36$°K and $a = 0.011$°K/mm Hg.

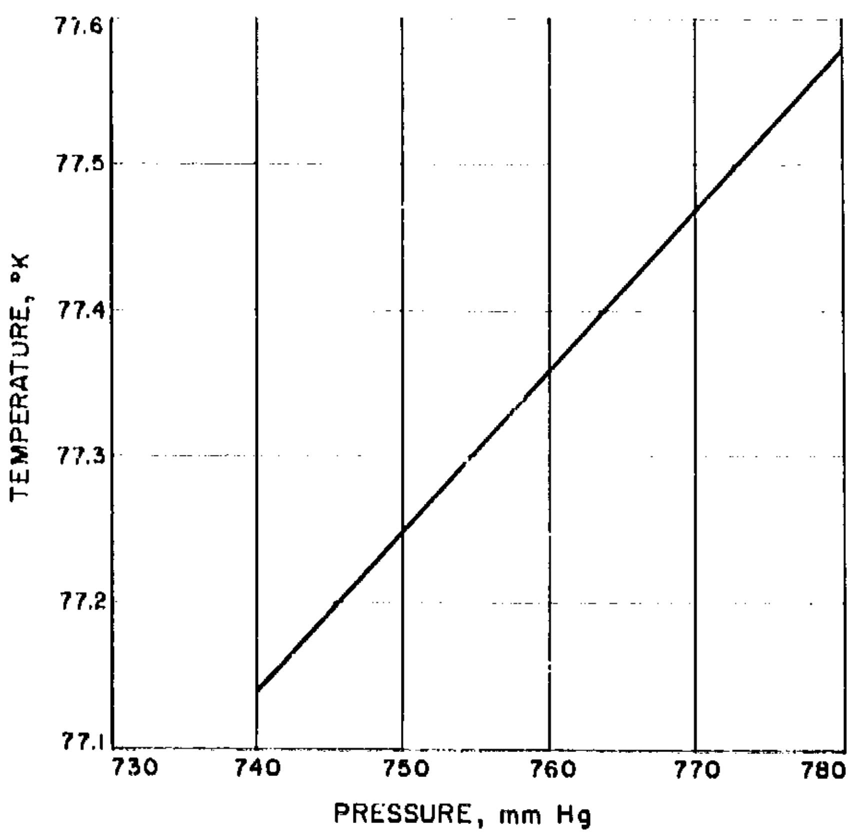


Fig. 6. Liquid nitrogen temperature versus vapor pressure

Substituting Eq. (10) into (6) and dropping second order terms,

$$
\begin{aligned}
T' \simeq\ & T(760) + a\Delta P \\
& + 0.23026\,[T_0 - T(760)]\left[\frac{L_1}{2}(\mathrm{db}) + L_2(\mathrm{db})\right] \\
& \cdot \left(1 + \frac{\Delta f}{f_0}\left\{\frac{\left(\frac{f_0}{f_c}\right)^4 - 6\left(\frac{f_0}{f_c}\right)^2 + 1}{\partial\left[\left(\frac{f_0}{f_c}\right)^4 - 1\right]}\right\}\right)
\end{aligned}
\qquad (11)
$$

B. Optical Communications Components

W. H. Wells

Various types of lasers and masers are providing oscillators for the infrared-microwave gap in the electromagnetic spectrum. This raises the question of whether little-known atmospheric transmission windows are suitable in clear weather for space experiments. Stations would have to be situated at sufficient altitude to escape most of the water vapor. Previous (*SPS 37-25*, Vol. IV, p. 122) crude estimates identified some possibilities. Some Russian theoretical work (Ref. 9) has been used to refine these estimates somewhat. Interpreted for our purposes, results show that only the windows at 10, 12, 18, 20, 860 and possibly 24 μ are likely to be useful at reasonable mountain altitudes, e.g., 11,000 ft at Jungfraujoch, Switzerland.

Far infrared lasers have been reported, out to at least 133 μ (0.133 mm). So far these are very inefficient devices, but there is potential for improvement. The atmosphere is opaque throughout most of the far infrared. Almost all of the absorption is due to water vapor, since it is the only asymmetric molecule of significant concentration in the atmosphere. However, water is concentrated in the lower altitudes, and a mountain site above 6000 ft would leave most of it below. The effects of the upper atmosphere are difficult to take into account, because most absorption measurements are made over horizontal laboratory paths.

Zhevakin and Naumov (Ref. 9) calculated the absorption spectrum of water vapor under conditions that approximate 10- and 20-km altitude. They also calculated absorption at certain selected wavelengths for sea level conditions. At 20 km transmission windows are abundant, but even at 10 km (32,808 ft), which can be reached only by aircraft, the water absorption spectrum is very restrictive. No attempt has been made to repeat the calculations for accessible mountain altitudes because the procedure is very involved, and because actual absorption depends upon humidity, temperature, and other trace gases. Therefore, only the published results were used for interpolation between the three altitudes 0, 10, and 20 km.

In the case of the transmission windows on the optical side of the infrared-millimeter wave gap, the three published absorption values fit an exponential curve and were readily integrated to give total absorption for one-way vertical transmission through the atmosphere. Table 1 gives the results in the form of parameters A and H for the equation

$$\text{db loss} = A\exp(-h/H). \qquad (1)$$

This table also lists the 1-db height (for vertical transmission) and wave numbers.

The three published values for each of the windows on the submillimeter side of the gap did not fit an exponential, so a crude graphical integration was used to estimate the height for 0.1-, 0.5-, and 1.0-db attenuation

Table 1. Total absorption for one-way vertical transmission through the atmosphere in some middle-infrared absorption windows

Wavelength, μ	Wave number, cm^{-1}	A for Eq. (1), db	H for Eq. (1) 1000 ft	1-db altitude, 1000 ft
10.4	966	0.77	4.2	1.1
11.6	860	1.0	4.3	0.06
17.9	560	3.6	4.3	5.5
20.1	497	6.9	4.5	8.6
24.4	410	16	4.5	13
27.4	365	33	4.6	16
31.4	318.5	68	4.9	21

(vertically). The results are given in Table 2. A small amount of absorption is more detrimental on this side of the spectrum, because it may contribute a large amount of thermal noise to a receiver that might otherwise have a noise temperature in the cryogenic range. Quantum rather than thermal noise dominates the optical side.

Previous results (*SPS 37-25*, Vol. IV) were expressed in the form of decibel attenuation above 11,000 ft (altitude of Jungfraujoch) at 50-deg zenith angle. For comparison, Table 3 gives the new results in this form for the three windows which are common to both sets of estimates. We have no explanation for the discrepancy

Table 2. Altitudes for 0.1-, 0.5-, and 1.0-db absorption in vertical transmission through the atmosphere at some submillimeter absorption windows

Wave-length, μ	Wave number, cm^{-1}	Frequency, Gc/sec	Altitude (1000 ft) at which vertical absorption is:		
			0.1 db	0.5 db	1.0 db
860	11.6	348	16	7.5	4.3
460	22	656	30	18	14
360	28	834	31	19	15
294	34	1020	47	31	25
200	50	1500	42	28	23

Table 3. Comparison with previous results: absorption at 50-deg zenith angle from Jungfraujoch, Switzerland

Wavelength, μ	New value	Old value
360	3	3
294	7	15
200	10	11

at 294 μ. Clearly, it was a mistake to consider the old estimates as upper limits of absorption. This was caused by neglecting a low pressure component of humidity that extends to great altitudes and contributes significant absorption despite its very low partial pressure. This component has a scale height of 4.2 km in contrast with 2.1 km for the main component [Ref. 9, Eq. (6)].

In conclusion, it appears likely that the windows at 10, 12, 18, 20, 860, and possibly 24 μ are sufficiently transparent to use from 11,000 ft. One might use other windows and tolerate high absorption in special types of spectroscopic experiments.

C. Antennas for Space Communications

A. Ludwig

1. Antenna Surface Measurement

A survey of the 30-ft antenna at the Venus site was made in January 1965 using a dual-transit method previously described (*SPS 37-31*, Vol. IV, p. 287). In addition to obtaining information about the antenna, further evaluation of the technique was accomplished. In this report, emphasis is on information pertinent to the technique itself, particularly concerning accuracy. It is concluded that the dual-transit technique is a useful tool in evaluating the performance of a high-frequency paraboloidal antenna.

a. Experiment description. The 30-ft antenna was nominally set at the RF boresight coordinates (of a 1964 K_μ-band experiment) of the 4-ft dish on the Echo tower on Tiefort Mountain. The 4-ft dish was read as a data point, thereby obtaining its x, y, and z coordinates in the same system as the rest of the data points. The location of this point relative to the axis of the best-fit and nominal paraboloids yields information on the RF boresight error. Forty points were read on the surface of the dish, and also four points were read to establish a reference plane with a known relationship to the antenna structure. The two theodolites were set approximately 60 ft apart, 30 ft in front of the dish. Two complete sets of data were obtained, resulting in improved accuracy and making possible a set-to-set comparison of results.

During the experiment it was discovered that the entire dish moved as a function of the position of the Sun. This was substantiated by theodolite readings and also by the antenna angle encoder readout. Fortunately the movement was confined primarily to a 1½-hr interval. No data were taken on the surface during this time, and by rereading the reference targets, corrections were obtained.

b. Accuracy of results. After reduction of the raw data by Dr. Charles Lawson's computer program, the various sets of data are best-fitted to one another by minimizing

$$\sum_{\text{All Points}} \Delta r^2 = \sum_{\text{All Points}} (\Delta x^2 + \Delta y^2 + \Delta z^2)$$

where differences are between a particular day's data and the average of two or more sets of data on a given point. Best-fitting includes both translation and rotation. Measurement errors are shown in Table 4. $\Delta\phi$ is the average of the absolute values of angular differences between lines-of-sight of the two transits, based on output from Dr. Lawson's program. Δr_1 is the physical error predicted for this $\Delta\phi$ based on the error analysis described in *SPS 37-31*, Vol. IV, p. 237. σ_x, σ_y, and σ_z are the standard deviations of Δx, Δy, and Δz after best-fitting as described above. Δr_2 is the average absolute magnitude of the error, derived statistically from σ_x, σ_y, and σ_z. Data from an earlier experiment (January 6) are included for reference, and are averaged in for determination of σ's on that date only. The large value of σ_x on that date is due to scaling or "stretching" relative to the other data. It is seen that Δr_2 compares well with Δr_1. This is important if only one set of data is taken, since Δr_2 could not be calculated in this case, and error estimates would have to be based on Δr_1 alone. Accuracy in a given direction is approximately ± 0.010 in. for a given set, which is excellent for this system. For the average value of two sets, σ_x, σ_y, or σ_z is expected to be approximately ± 0.007 in.

In the reference system, the coordinates of the Echo tower 4-ft dish on Tiefort Mountain are typically (several observations were made) $x = 441.7$ in., $y = 256.1$ in., $z = 803{,}992.9$ in. This range of 66,999 ft is 670 ft greater than the value determined by radar measurement. This corresponds to an angular error of 1.9 sec of arc at that distance, with the 60-ft baseline used.

c. Best-fit paraboloid and RF pointing error. The average values of the two sets of complete data, corrected to the reference plane coordinate system, were input to the Utku-Barondess best-fit paraboloid computer program (Ref. 10). The results are shown in Tables 5 and 6, compared to data obtained by Rohr Corp. in the same month and also to the deflections calculated by the STAIR program. All three sets of data use the same best-fitting program. Agreement is generally good. Due to a different distribution of data points, some disagreement is expected. The point-by-point deflections from the best-fit paraboloid are shown in Fig. 7.

The RF boresight was found to be pointing high (+0.011 deg) and to the right (+0.031 deg) of the perpendicular to the reference plane. Calculated RF boresight error, using the data shown in Table 5 and additional data on hyperbola tilt and translation, is compared to

Table 4. Dual transit system measurement accuracy

Data of:	$\Delta\phi$, sec	Δr_1, in.	σ_x, in.	σ_y, in.	σ_z, in.	Δr_2, in.
1/6/65	—	—	0.040	0.017	0.010	0.038
1/26/65	±4.5	0.024	0.008	0.011	0.014	0.019
1/27/65	±2.6	0.014	0.008	0.007	0.009	0.014
1/28/65	±3.1	0.017	0.008	0.007	0.009	0.014

Table 6. Surface RMS deviation before and after best-fit paraboloid

Source	Before	After
Dual-transit	0.038	0.024
Vertex-transit	0.042[a]	0.023[a]
Calculated (STAIR)	0.063[a]	0.013[a]

[a] Manufacturing tolerance on panels has not been included. This reflects deflections of the panel tie-points only.

Table 5. Characteristics of best-fit paraboloid

Source	Coordinates of vertex			Coordinates of focus			Rotation of z-axis	
	x	y	z	x	y	z	α_x, deg	α_y, deg
Dual-transit	0.304	0.619	0.027	0.189	0.317	150.023	0.116	0.044
Vertex-transit	0.143	0.395	0.027	0.115	0.191	150.009	0.078	0.011
Calculated (STAIR)	0.000	0.919	0.021	0.000	0.492	150.020	0.163	0.000

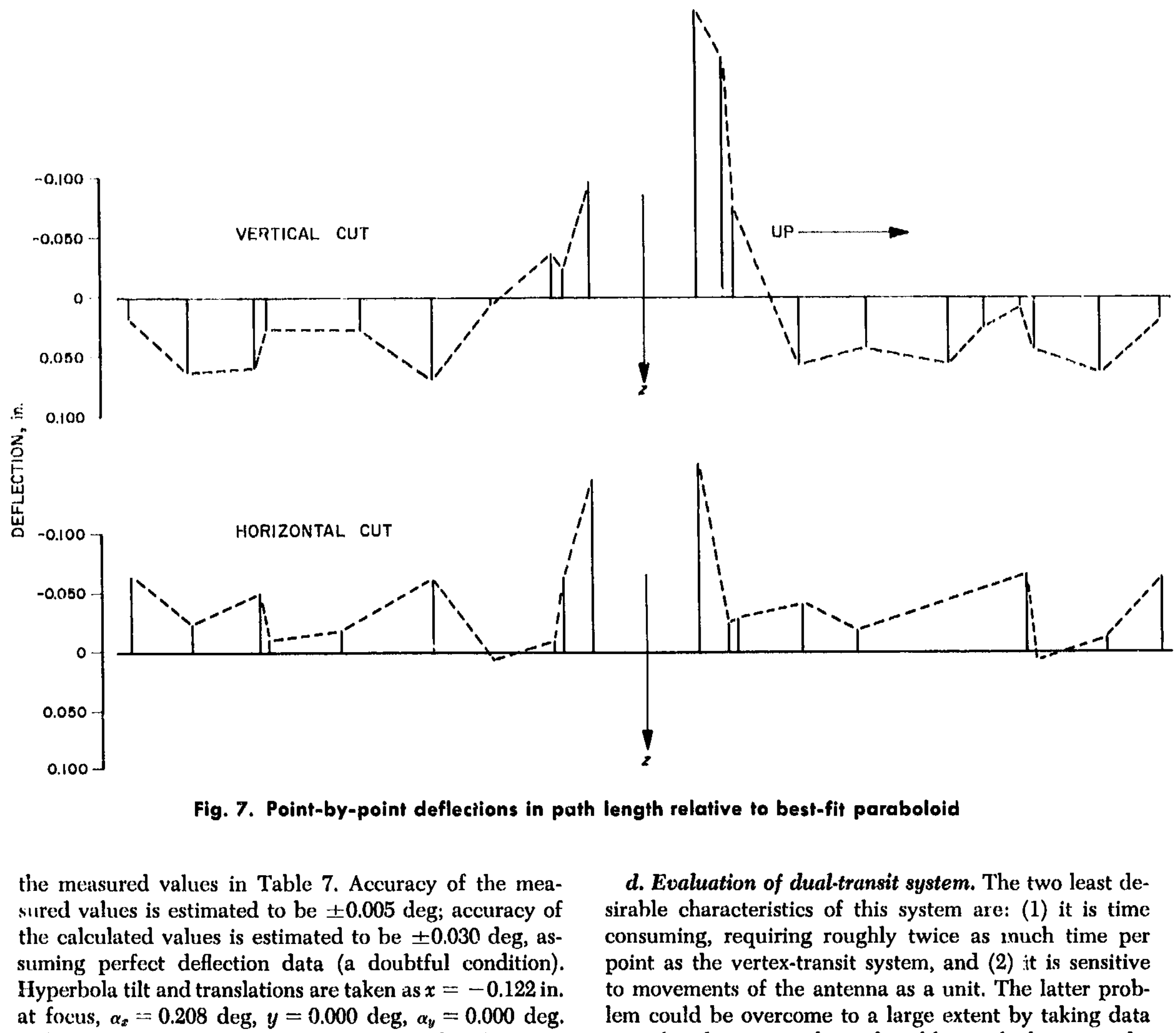


Fig. 7. Point-by-point deflections in path length relative to best-fit paraboloid

the measured values in Table 7. Accuracy of the measured values is estimated to be ± 0.005 deg; accuracy of the calculated values is estimated to be ± 0.030 deg, assuming perfect deflection data (a doubtful condition). Hyperbola tilt and translations are taken as $x = -0.122$ in. at focus, $\alpha_x = 0.208$ deg, $y = 0.000$ deg, $\alpha_y = 0.000$ deg. A beam factor of 0.8 was used. It is possible, by combining the data in Tables 5 and 7 to determine pointing relative to the best-fit paraboloid axis, or other interesting reference systems.

Table 7. RF boresight error, 1964 K_u-band experiment

Source	Azimuth, deg	Elevation, deg
Dual-transit direct measurement	+0.031	+0.011
Dual-transit calculated	+0.022	−0.009
Vertex-transit calculated	−0.015	−0.010
STAIR program calculated	0.000	−0.003

d. Evaluation of dual-transit system. The two least desirable characteristics of this system are: (1) it is time consuming, requiring roughly twice as much time per point as the vertex-transit system, and (2) it is sensitive to movements of the antenna as a unit. The latter problem could be overcome to a large extent by taking data at night, eliminating thermal problems which seem to be the most troublesome. The accuracy obtained in this experiment was quite good. This is due in no small part to the exceptional care taken by the transit operators and also to the point-by-point check for consistency between forward and reverse readings as the data were taken. A "production" run would probably not repeat these high standards. Generally speaking, the method appears to be competitive regarding accuracy of the surface measurement. The built-in error estimate appears to work, and is a valuable extra benefit. This system is admirably suited for determining antenna pointing (it is felt that the measurements on pointing are better than any prior experiment) and could be used for this purpose alone.

References

1. Sees, J. E., "Fundamentals in Noise Source Calibrations at Microwave Frequencies," NRL Report 505, p. 5, Eq. (14), Radio Astronomy Branch, Atmosphere and Astrophysics Division, Naval Research Laboratory, Washington, D.C., January 1958.
2. Worthing, A. G., and Geffner, J., "Treatment of Experimental Data," John Wiley & Sons, Inc., 9th printing, p. 209, Eq. (32), New York, 1960.
3. Tiuri, M. E., "Radio Astronomy Receivers," *IEEE Transaction on Military Electronics*, Vol. MIL-B Nos. 3 and 4, p. 267, July–October 1964.
4. Ramo, S. and Whinnery, J. R., "Fields and Waves in Modern Radio," p. 44, John Wiley and Sons, Inc., Second Edition, 1953.
5. Stelzried, C. T., "Temperature Calibration of Microwave Thermal Noise Sources," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-13 No. 1, p. 128, January 1965.
6. Oakley, C. O., "The Calculus," p. 192, Barnes and Noble Inc., New York, 1944.
7. Reference Data for Radio Engineers, Fourth Edition, p. 627, International Telephone and Telegraph Corp., New York.
8. White, G. K. "Experimental Techniques in Low-Temperature Physics," Clarendon Press, Table XI, p. 104, 1959.
9. Naumov, A. P. and Zhevakin, S. A. "Absorption of Electromagnetic Radiation by Water Vapor on 10 μ – 2 cm Waves in the Upper Layers of the Atmosphere," Geomagnetism and Aeronomy, Vol. 3, No. 4, pp. 537–46, 1963.
10. Barondess, S. M., and Utku, S., "Computation of Weighted Root Mean Square of Path Length Changes Caused by the Deformations and Imperfections of Rotational Paraboloidal Antennas," Technical Memorandum No. 33-118, Jet Propulsion Laboratory, Pasadena, California, March 1, 1963.

XXIV. Communications Systems Research: Information Processing

A. Rejection of Outliers in Radar Observations

E. C. Posner

This article presents a criterion for rejecting outlying radar observations caused by a measurement error. A particular error model is assumed in which with a small probability there is an offset in the true range added on to random errors of observation. Maximum-likelihood techniques then give a reasonable rejection criterion, which is of the form "reject when one observation differs from the mean by a certain constant times the standard deviation of the measurements."

1. Introduction

Many estimation problems in radar reduce themselves to the following form. There are K independent normal random variables (or observations) of unknown mean μ and known variance σ^2. The parameter μ to be estimated is the unknown range. The variance σ is assumed known. In practice each observation is an average of many, so that the variance can be estimated; or, alternatively, the σ^2 may be known because the receiver signal-to-noise ratio is known.

The error model is as follows: these are K observations y_i, $1 \leq i \leq K$. With probability $1 - \epsilon$, ϵ small but known, y_i is chosen from the normal distribution of mean μ, variance σ^2 [i.e., y_i is $N(\mu, \sigma^2)$]. With probability ϵ, y_i is chosen from $N(\mu + \Delta, \sigma^2)$, where Δ is the unknown offset. Such an i is called an error position.

The occurrence of an error at position i is assumed to be independent of an error at j, $j \neq i$. This is reasonable since the observations are taken on different nights by different observers, and errors are introduced by missetting dials, misreading ephemerides, etc., during different observations.

2. Likelihood Ratios

We now treat the problem of finding the maximum likelihood estimate of the mean μ for the model presented in the preceding section. We incidentally arrive at the maximum likelihood estimate of the offset Δ.

We said that ϵ is assumed small, but now we must be more specific. Namely, ϵ is so small that an occurrence of two or more errors in the K observations is very unlikely even compared with the probability of a single error. Let us demand that the probability of a double error, given a single error, be less than 0.01. If $K = 40$, how small must ϵ be? The probability of another error is $(K - 1)\,\epsilon(1 - \epsilon)^{K-2} \sim (K - 1)\epsilon$ for ϵ small. Hence,

$$\epsilon \leq \frac{0.01}{39} = 2.56 \times 10^{-4}.$$

It will turn out that the maximum-likelihood procedure for estimating μ is to first locate the error position and then reject the observation. We then estimate μ from the remaining positions in the usual manner. Thus, even if it is suspected that two errors have occurred, the first error position could be rejected. The procedure could then be repeated with one less observation. Therefore, a restriction on ϵ is not so serious.

Once μ is estimated, Δ is estimated as the difference between the rejected observation and the estimate of μ. Thus, the problem reduces to estimating μ. To do this, we develop the techniques which determine the best procedure.

The likelihood function of the K observations y_i must first be written down. Define a random variable $w(i)$, $1 \leq i \leq K$, to be 1 if the ith position is an error position, and zero otherwise. Then the w_i are independent, by assumption. Consequently, the likelihood of the observed K-vector $y = (y_1, \cdots, y_K)$ is given by

$$L(y;\mu,\Delta) = (1-\epsilon)^{K - \sum_{i=1}^{K} w_i}\,\epsilon^{\sum_{i=1}^{K} w_i}\,\frac{1}{\sigma^K}\prod_{i=1}^{K}\frac{1}{(2\pi)^{1/2}}$$
$$\exp\left[\frac{-1}{2\sigma^2}\left\{(y_i - u - w_i\Delta)^2\right\}\right]. \qquad (1)$$

Since ϵ is assumed small, we assume that at most one error has been made. We first make a maximum-likelihood estimate of j, the unique error position. It can be seen, with or without reference to Eq. (1), that the maximum-likelihood estimate of the error position j, given that exactly one error occurred, is that j which maximizes

$$\left| y_j - \frac{1}{K}\sum_{i=1}^{K} y_i \right|.$$

That is, the y_j farthest away from the mean of all the y_i is found. This w_j is 1. The other w_i are 0, it only one error occurred. This j in any case will be called the outlying position.

The philosophy is to maximize the value of L in Eq. 1 under each of two assumptions: $w_j = 1$ and $w_j = 0$. The assumption yielding the larger likelihood is assumed correct. If it turns out that $w_j = 1$ is assumed, then an outlying observation is assumed at j. If $w_j = 0$ is assumed, no outliers are assumed to have occurred.

The index j as candidate for error position having been determined, the log likelihood, if an error was indeed made at j, is given by $\log L_1$, where

$$\log L_1 = (K - 1)\log(1 - \epsilon) + \log\epsilon - K\log(2\pi)^{1/2} - K\log\sigma$$
$$- \frac{1}{2\sigma^2}\sum_{\substack{i=1\\ i\neq j}}^{K}(y_i - \mu)^2 - \frac{1}{2\sigma^2}(y_j - \mu - \Delta)^2. \qquad (2)$$

And if there were no error, the log likelihood $\log L_2$ becomes

$$\log L_2 = K\log(1 - \epsilon) - K\log(2\pi)^{1/2} - K\log\sigma$$
$$- \frac{1}{2\sigma^2}\sum_{i=1}^{K}(y_i - \mu)^2. \qquad (3)$$

The object is now to maximize $\log L_1$ by choice of μ and Δ, and to maximize $\log L_2$ by choice of μ. Whichever maximization gives the larger maximum determines the final maximum-likelihood estimator of μ.

3. *Maximizing the Two Likelihoods*

We first maximize $\log L_1$, the likelihood of the K observations under the assumption that j, the outlying position, was in fact an error position. In Eq. (2) there are two parameters μ and Δ. To maximize, we first take partial derivatives with respect to μ and Δ. Thus, we need

$$\frac{\partial \log L_1}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{K}(y_i - \mu) - \frac{\Delta}{\sigma^2}; \qquad (4)$$

$$\frac{\partial \log L_1}{\partial \Delta} = \frac{1}{\sigma^2}(y_j - \mu - \Delta). \qquad (5)$$

Let us denote the solutions of

$$\frac{\partial \log L_1}{\partial \mu} = \frac{\partial \log L_1}{\partial \Delta} = 0$$

by μ_1 and Δ_1, the maximum-likelihood estimators of μ and Δ given that an error really did occur at the outlying position j. From Eq. (5), we have

$$\Delta_1 = y_j - \mu_1. \qquad (6)$$

Using Eq. (6), Eq. (4) yields

$$\sum_{i=1}^{K}(y_i - \mu_1) = y_j - \mu_1. \qquad (7)$$

But Eq. (7) can be written

$$\mu_1 = \frac{1}{K-1}\sum_{i \neq j} y_i . \qquad (8)$$

That is, if an error really occurred at j, the maximum-likelihood estimate of μ is obtained by rejecting the outlying observation and then estimating μ in the usual way with the remaining $K - 1$ observations. We have therefore proved one of the results described in part 3.

Now let us maximize L_2, the likelihood under the assumption that no errors occurred. This is then the standard problem of estimating the mean of a normal distribution, and as usual

$$\mu_2 = \frac{1}{K}\sum_{\text{all } i} y_i . \qquad (9)$$

4. The Final Maximum-Likelihood Estimator

The next step is to substitute the conditional maximum-likelihood estimators into Eqs. (2) and (3), and decide which maximum is the larger. If L_1 is larger, assume that an error occurred at j; if L_2 is larger, no error occurred. The final maximum-likelihood estimator $\hat{\mu}$ of μ is then either μ_1 or μ_2, depending upon whether L_1 or L_2 turns out larger. We actually base our decision on the sign of $\log L_1 - \log L_2$ when μ_1, Δ_1, are substituted for μ, Δ in L_1, and μ_2 substituted for μ in L_2. The resulting difference is called Λ.

Using Eqs. (6), (8), and (9), we find, after the convenient location change making $\mu_2 = 0$,

$$\Lambda = \log\epsilon - \log(1-\epsilon) + \frac{K}{K-1}\,\frac{y_j^2}{2\sigma^2}. \qquad (10)$$

Eq. (10) is the equation for Λ, where if $\Lambda > 0$, assume an error occurred and use μ_1, and if $\Lambda < 0$, assume that an error did not occur and use μ_2. Recall that in Eq. (10), $\sum_{i=1}^{K} y_i$ is defined as 0, however.

5. Criterion

Eq. (10) finally allows us to obtain the maximum-likelihood estimator of μ. We sum up the results of the preceding sections in the following criterion; $\bar{y}$ is defined as $\frac{1}{K}\sum_{i=1}^{K} y_i$.

First find the j that maximizes $|y_j - \bar{y}|$. Estimate μ as the mean of all K observations if

$$(y_j - \bar{y})^2 < \left\{\frac{K-1}{K}\right\} 2\sigma^2 \log\left(\frac{1-\epsilon}{\epsilon}\right). \qquad (11)$$

On the other hand, if

$$(y_j - \bar{y})^2 > \left\{\frac{K-1}{K}\right\} 2\sigma^2 \log\left(\frac{1-\epsilon}{\epsilon}\right), \qquad (12)$$

reject the j^{th} observation from the data. Then estimate μ as the mean of the remaining $K - 1$ observations.

What the criterion says is that if $|y_j - \bar{y}|$ is too large in comparison with σ, the j^{th} observation should be ignored. This is certainly a reasonable criterion and, in fact, one which many people use on intuitive grounds. The point of this article is that the definition of "too large" can be made quite precise from a reasonable model of the situation.

Let ρ be the ratio of $|y_j - \bar{y}|$ to the known σ, above which ratio rejection occurs. We have proved that

$$\rho = \left[2\left(\frac{K-1}{K}\right)\log\left(\frac{1-\epsilon}{\epsilon}\right)\right]^{1/2}. \qquad (13)$$

A good feature of Eq. (13) is that ρ does not depend very much on ϵ, since the dependence is as $(\log 1/\epsilon)^{1/2}$ for ϵ small. Thus, only approximate knowledge of ϵ allows outliers to be rejected. Assuming $\epsilon = 2 \times 10^{-4}$ and $K = 40$, one has $\rho = 4.08$. This can be called a "4σ" criterion. The probability that a given observation in the absence of an error is more than 4σ away from the mean is 3.2×10^{-5}. The rigorous maximum-likelihood criterion in this case turns out to reject fewer outliers than intuitive criteria, which often are "3σ" criteria. But of course all this depends upon how small ϵ is assumed to be *a priori*.

B. Identification of Finite State Machines

E. C. Posner

This article introduces an area of study ultimately motivated by the problem of fault identification in digital machines, one of the main problems facing the DSN in digital control and communication. It is proved that given a finite class of convivial irredundant nonisomorphic binary machines in an unknown state, there is an input sequence of finite length such that the output sequence uniquely determines the machine and the final state. The results are applied to the problem of identifying miswiring in ranging coders.

1. Introduction

We begin with a brief summary of definitions used in the theory of finite state machines. A finite state machine M is defined as an ordered five-tuple $M = (K, A, B, \mu, \delta)$ having the following properties:

(1) K is a finite set, called the *state space*.

(2) A is a finite set, called the *input alphabet*.

(3) B is a finite set, called the *output alphabet* (A and B can be the same set).

(4) μ is a function from $A \times K \to B$, called the *output function*.

(5) δ is a function from $A \times K \to K$, called the *next-state function*.

What all this means is that the machine accepts inputs, and produces an output and a next state depending only on the input and the present state. Two special definitions needed for convenience to rule out pathological counterexamples are the following:

Definition: A finite state machine M is said to be *convivial* if and only if for every pair of states K_1 and K_2 in K, there is a sequence of inputs $\{a_i, 1 \leq i \leq n(K_1, K_2)\}$ such that if the machine is in state K_1 at time 0, it is in state K_2 at time n.

The point of the above definition is that the all states communicate with each other by suitable inputs. Note that although "time" is not explicitly called out in the definition of a finite state machine, it is implicit in the definition. We envision an input sequence $a_1, a_2, \cdots, a_n$, and call the time before a_1 "arrives" time zero. The outputs $b_1, b_2, \cdots, b_n$ are then time sequenced in a similar way.

A second definition needed is that of *irredundancy*.

Definition. A finite state machine M is called *irredundant* if and only if given two different states K_1 and K_2 of K, there is an input sequence $\{a_i, 1 \leq i \leq n\}$ for M such that the output sequences $\{b_i, 1 \leq i \leq n\}$ are different for K_1 and K_2.

The meaning of this definition is just that if two states always give the same output sequence for every input sequence, they are the same state. Without such a definition, there would be no hope of ever identifying the state a machine is in.

We also need to define an isomorphism between two machines.

Definition. Two finite-state machines M_1 and M_2 are called *isomorphic* if and only if there are one-one correspondences $f\colon K_{(1)} \to K_{(2)}$, $K_{(1)}$, $K_{(2)}$, the respective state spaces, $g\colon A_{(1)} \to A_{(2)}$ $A_{(1)}$, $A_{(2)}$, the respective input alphabets, and $h\colon B_{(1)} \to B_{(2)}$, $B_{(1)}$, $B_{(2)}$, the respective output alphabets, such that, if $\mu_{(1)}$, $\mu_{(2)}$ are the respective output functions, and $\delta_{(1)}$, $\delta_{(2)}$ the respective next-state functions, then

$$\delta_{(1)} = h^{-1} \cdot \delta_{(2)} \cdot (g \times f)\,; \tag{1}$$

$$\mu_{(1)} = f^{-1} \cdot \mu_{(2)} \cdot (h \times f)\,. \tag{2}$$

The meaning of this definition is simply that two machines are isomorphic if they behave the same under relabelling of alphabets and states. The triple f, g, h, is called the *isomorphism*. In case $A_{(1)} = A_{(2)}$, $B_{(1)} = B_{(2)}$, we define a stronger concept (fewer things isomorphic) called *state isomorphic*. The definition of state isomorphism is the same as isomorphism, except that g and h are to be the identity mapping. We then call the mapping $f\colon K_{(1)} \to K_{(2)}$ the isomorphism. We remark that irredundancy can also be defined in terms of state isomorphisms.

In this note, we are concerned with *binary machines*. A binary machine is one in which A and B have two elements. We then identify A and B with the set $\{0, 1\}$. It can be shown that there is no loss of generality in considering only binary machines, since fictitious states can be invented in M. Details are left to the reader. A good general background in this subject is Moore's recent book (Ref. 1).

2. Main Theorem

In this section, we present a theorem on the identification of binary irredundant convivial finite state machines. The result is not deep, but finding the best bounds for length of input sequence necessary for special classes of finite state machines is highly nontrivial.

Theorem. Let F be a known finite class of binary irredundant convivial finite state machines $M_1, M_2, \cdots, M_r$ initially in states $K_0^{(1)}, K_0^{(2)}, \cdots, K_0^{(r)}$, respectively. Let no two machines in F be state isomorphic. Then there exists a positive integer $n = n(F)$ and an input sequence $\alpha = \{a_i, 1 \leq i \leq n\}$ such that, by observing the output sequence $\beta = \{b_i, 1 \leq i \leq n\}$, a machine known to be in F can be identified, together with its final state at time n.

Proof. We first produce an integer n and an input sequence of length n with the property that the output sequences of length n corresponding to every machine in F (in any possible initial state) are all different. One then observes the output of length n and looks up in a prepared table the unique machine that can produce such an output sequence of length n. The construction will be such that the state of the machine found at time n can be determined.

A difficulty in practice, in addition to finding the input sequence with this property, is to find a good way of going from output sequence to unknown machine. Thus, this proof is essentially nonconstructive (although not in a strict metamathematical sense), and the work necessary to find these inputs and correspondences may prove to be very difficult.

To produce such an input sequence, we reason as follows. List all the ordered pairs of machines and F-initial states, and call this list $L = \{L_1, L_2, \cdots, L_t\}$. By irredundancy and by the fact that no two machines in F are state isomorphic, every two elements of L can be distinguished by a sufficiently long input sequence. Let α_1, of length n_1, be an input sequence which distinguishes L_1 from L_2. That is, the output under the hypothesis that the original machine-state pair is L_1 is different from the output under the L_2 assumption.

Of course the output after the sequence α_1 of length n_1 is inputted may not be a possible output under either L_1 or L_2, but in any case, at least one L_i is ruled out. Furthermore, the input sequence α_1 causes a known transformation (not necessarily one-to-one or onto) of the list L, so that after α_1 we are confronted with a new shorter list L'.

We must prove that no two machine states of L' coming from different machines can yield the same output for every input. But this possibility is readily ruled out by conviviality and the fact that no two machines of F are state-isomorphic. Thus, the identity of the original machine does not get lost in this reduction process

Continuing in this way, we arrive with longer and longer input sequences at a list L^* consisting of a unique machine-state pair. This proves the theorem.

We remark that in general it can be shown to be impossible to identify the original state of the machine that has been found. But by conviviality, the found machine can now be put into any state desired by a known input sequence. And for special classes F of machines, even the original state can be identified (for example, if the class consists solely of *binary linear shift registers*).

3. Application to Shift Registers

We now apply the main theorem to special classes F, getting constructive input sequences and constructive mappings of output sequences to machine-state pairs.

In particular, let F consist solely of *linear shift registers* (Ref. 2). The output sequence β is given by $\alpha = p(D)\beta$, where $p(D)$ is the *shift register polynomial* with mod-2 coefficients in the *difference operator* D. For example, in Fig. 1, $n = 7$, $p(D) = D^7 \oplus D^6 \oplus D^3 \oplus 1$ and $a_k = b_k \oplus b_{k-2} \oplus b_{k-4} \oplus b_{k-7}$. That is, $b_k = a_k \oplus b_{k-2} \oplus b_{k-4} \oplus b_{k-7}$, $k \geq 0$ ($b_{-1}, b_{-2}, \cdots, b_{-7}$ are the *initial condition* of the register). Linear shift registers form the heart of the JPL ranging systems (Ref. 3).

As a finite state machine, a shift register has 2^n states, corresponding to the 2^n n-tuples of 0 and 1. The output and next-state functions are readily defineable from $p(D)$. For classes of linear shift registers, one readily also verifies conviviality, irredundancy, and non-state isomorphism. Furthermore, two linear shift registers are state-isomorphic if and only if their association polynomials are the same. The theorem of the preceding section then applies, showing that given a finite family of linear shift registers, one can identify which register it is by observing the output of a fixed finite input sequence. In fact, the initial state can even be identified, since one can prove that no input causes two distinct states to collapse into one state.

Consider the following application: suppose a shift register $p(D)$ is to be wired up. It is discovered that one stage (or flip-flop) has not been wired in. Find out where to wire it.

We must first find the class F. Each register in F has $n - 1$ instead of n stages, and there are as many machines in F as there are runs of 0's in 1's in the ordered set of coefficients of $p(D) \oplus D^n$. (A *run* in a sequence is a maximal consecutive set of the same symbol). For, removing a stage from a run yields the same polynomial as removing any other stage from the same run; conversely, different runs yield different polynomials. For example, if $p(D) = D^7 \oplus D^5 \oplus D^3 \oplus 1$, then F has 6 elements, corresponding to the 6 polynomials in Table 1.

Table 1. The Class *F*

D^6	$\oplus$	D^5	$\oplus$	D^3	$\oplus$	1
D^6	$\oplus$	D^3	$\oplus$	1		
D^6	$\oplus$	D^4	$\oplus$	D^3	$\oplus$	1
D^6	$\oplus$	D^4	$\oplus$	1		
D^6	$\oplus$	D^4	$\oplus$	D^2	$\oplus$	1
D^6	$\oplus$	D^4	$\oplus$	D^2		

The polynomials were obtained by deleting the seven stages one at a time starting from the left in Fig. 1; there are exactly one run of length 2 and five runs of length 1. The last register $D^6 \oplus D^4 \oplus D^2$ does not have a D^0 term and is really a 4-stage register $D^4 \oplus D^2 \oplus 1$, but this does not affect the argument.

Now let us find a sequence which uniquely identifies which linear shift register in a finite class F we have. First we claim we can set the unknown register to the all zero state by a suitable input sequence. For, since the registers are known, their periods are known for every initial condition. Details are left to the reader.

So let all the elements of F be known to be the zero initial condition. For convenience, let us assume that all lengths of the registers in F are equal (the case in the application of miswiring); the argument works in general, however. Then input the sequence $100 \cdots 0$, and notice how long it takes for a 1 to appear at the output; stop the sequence of 0's at this point. This gives the first nonzero term after the D^n term in $p(D)$. Then follow this sequence with another $10 \cdots 0$. The output at each time is 0 or 1 in a known way depending on whether the original 1 that was entered has "met" the second nonzero summand of $p(D)$ beyond D^n. Continuing in this way, the entire polynomial is identified with a very short input sequence, much shorter than that given by the general theorem. Furthermore, the correspondence between outputs and machines is natural in this case, but not in the general case.

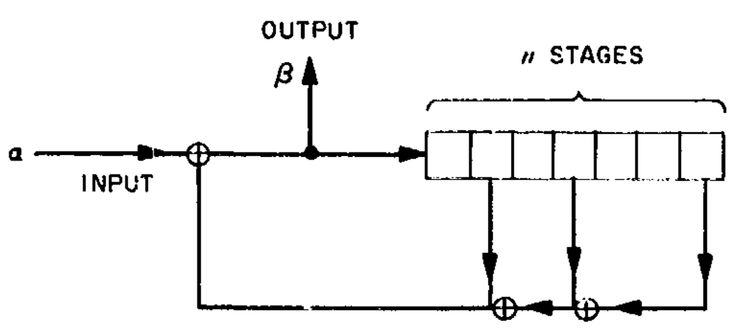


Fig. 1. Linear shift register

For example, suppose the register $p(D) = D^3 \oplus D \oplus 1$ has been miswired, resulting in an F of 2 elements: $D^2 \oplus D \oplus 1$ and $D^2 \oplus 1$. Assuming the initial condition 00 in the register, input a 1. If the output is a zero, the machine is $D^2 \oplus 1$; if the output is a 1, the machine is $D^2 \oplus D \oplus 1$. (The verification is left to the reader.) Thus, an input sequence of length 1 works here!

Better and more specific results on fault identification of digital machines would be extremely useful in maintaining a low down time for the digital equipment in the DSN.

C. An Improved Decoding Procedure for Binary Bose-Chaudhuri Codes

E. Berlekamp[1]

This article presents a way of cutting down the computational time necessary to decode a Bose-Chaudhuri code over the two-symbol alphabet. The solution of Newton's identities over a field of characteristic two is the key step in decoding these codes in all decoding procedures. This article introduces a simple transformation of variables which halves the size of the resulting matrix which must be inverted. This transformation can increase the decoder's speed by almost a factor of eight and decrease the decoder's storage requirements by almost a factor of four.

1. Introduction

In order to decode a linear binary code one considers the set of all error patterns which have the same syndrome (pattern of parity check failures) as the received syndrome. From that set one chooses an error pattern

[1]University of California, Berkeley.

which has the fewest errors. A solution to this problem is a set of *error-location numbers* β_i. Bose-Chaudhuri (Refs. 4 and 5) have suggested that if the block length n is one less than a power of 2 then the positions of the code (and hence the error-location numbers) may be taken as the nonzero elements of a finite field $GF(n+1) = GF(2^k)$. Bose-Chaudhuri have demonstrated methods of choosing a parity check matrix for the code so that the syndrome of the received sequence may be interpreted as the first t odd power-sum symmetric functions,

$$S_j = \sum_i \beta_i^j ; j = 1, 3, \cdots, 2t-1.$$

The decoding problems for Bose-Chaudhuri codes consists of finding the β_i, given S_j, $j = 1, 3, \cdots, 2t-1$ [and implicitly given S_j for $j = 2, 4, \cdots 2t$ since $S_{2i} = S_i^2$ in $GF(2^k)$]. One form of a solution to the decoding problem is the polynomial $\pi_i (x - \beta_i) = \sum_{i=0}^{t} \sigma_{t-i}\, x^i$. Chien (Ref. 6) has recently demonstrated an elegant, easily instrumented method for finding all the roots in $GF(2^k)$ of any polynomial $\sigma(x)$, but the computation of the σ's from the S's remains a difficult task.

Let t be the number of errors which the code is designed to correct. (Combinatorial arguments reveal that in many cases the code must be capable of correcting substantially more errors than is guaranteed by this lower bound, but, unfortunately, no practical means of decoding any of the correctible error patterns containing more than t errors is yet known.) If there are no more than t errors, then for $j > t$, $\sigma_j = 0$, and for $j \leq t$, the relations between the σ's and S's are given by Newton's Identities (Ref. 7):

$$\begin{bmatrix} S_1 & 1 & 0 & 0 & \cdots & 0 \\ S_3 & S_2 & S_1 & 1 & \ddots & \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ S_{2t-3} & & & & & \\ S_{2t-1} & S_{2t-2} & \cdots & & & \end{bmatrix} \begin{bmatrix} 1 \\ \sigma_1 \\ \sigma_2 \\ \cdot \\ \cdot \\ \cdot \\ \sigma_t \end{bmatrix} = 0 \qquad (1)$$

The determinant of the $t \times t$ submatrix consisting of the last t columns of the above matrix is conventionally denoted by M_t. If $M_t \neq 0$, then the t in Eqs. (1) are independent; if $M_t = 0$, they are dependent.

One method of computing the determinant M_t is to triangularize the matrix in Eqs. (1), using a computer with galois field arithmetic units. This process requires about $t^3/3$ galois field operations, and about t^2 words of memory to store the matrix while it is being triangularized.

If $M_t \neq 0$, the decoder might complete the process of solving Eqs. (1) for the σ's and then find the error location numbers by the methods of Chien (Ref. 6). If $M = 0$, the decoder might assume that $\sigma_t = \sigma_{t+1} = 0$, and then attempt to solve a system of $t-1$ equations for σ_i, $i = 1, 2, \cdots t-1$. Various embellishments in this procedure have been proposed by Chien, Massey, and Solomon (Refs. 6, 8 and 9). By various clever methods, these recent schemes enable the decoder to compute the error location numbers without ever actually finding the σ's. However, all known decoding schemes still require calculations of M_t. Unless t is very small, this calculation remains the most difficult step in the decoding procedure. Since our transformation halves the size of this crucial determinant, it can be profitably incorporated into any of the decoding schemes that have been suggested. For some schemes, such as those suggested by Forney (Ref. 10) for decoding when both erasures and errors are present, the inclusion of our transformation may however necessitate substantial changes in the decoding procedure.

2. The Transformation

Suppose we introduce some new variables R_k by convolving the sequence S_k with some sequence A_k. Let

$$R_k = \sum_j A_j S_{k-2j}, \text{ where } A_0 = S_0 = 1;$$

$$A_{-i} = S_{-i} = 0 \text{ for } i > 0 \qquad (2)$$

Then if we add to each Eq. (1) A_1 times the previous equation, A_2 times the equation before that, $\cdots$, we transform (1) into

$$\begin{bmatrix} R_1 & 1 & 0 & 0 & \cdots & 0 \\ R_3 & R_2 & R_1 & 1 & \ddots & \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ R_{2t-1} & R_{2t-2} & \cdots & & & \end{bmatrix} \begin{bmatrix} 1 \\ \sigma_1 \\ \sigma_2 \\ \cdot \\ \cdot \\ \cdot \\ \sigma_t \end{bmatrix} = 0 \qquad (3)$$

More succinctly, $\sum_j R_{2i-j+1}\sigma_j = 0$ for $i = 1, 2, \cdots, t$. (4)

Notice that this transformation is readily invertible:

$$S_k = R_k + \sum_{j=1}^{k} A_j S_{k-2j}. \tag{5}$$

We now choose the A_j in such a way that $R_{2i} = 0$ for all $i \neq 0$. This is readily accomplished by setting

$$0 = \sum_{j=0}^{i} A_j S_{2(i-j)} \qquad \text{or } A_i = \sum_{j=0}^{i-1} A_j S_{2(i-j)}. \tag{6}$$

Given the initial condition that $A_0 = 1$, Eq. (6) defines the rest of the A_i recursively.

With $R_{2i} = 0$, Eq. (3) can be separated into two systems of equations. The last $[t/2] + 1$ Eqs. from (3) becomes Eqs. (7); the first $t - [t/2]$ Eqs. becomes Eqs. (8).

$$\begin{bmatrix} \ddots & & & \ddots \\ & \ddots & \ddots & \\ & & \ddots & \ddots \\ R_{2t-5} & & & \ddots \\ R_{2t-3} & R_{2t-5} & & \ddots \\ R_{2t-1} & R_{2t-3} & R_{2t-5} & \end{bmatrix} \begin{bmatrix} 1 \\ \sigma_2 \\ \sigma_4 \\ \cdot \\ \cdot \\ \cdot \\ \sigma_{2\left[\frac{t}{2}\right]} \end{bmatrix} = 0; \tag{7}$$

$$\begin{bmatrix} \sigma_1 \\ \sigma_3 \\ \sigma_5 \\ \cdot \\ \cdot \\ \cdot \\ \\ \end{bmatrix} \begin{bmatrix} R_1 & 0 & 0 & \cdots & 0 \\ R_3 & R_1 & 0 & & 0 \\ R_5 & R_3 & R_1 & \ddots & \\ & & & & \\ & & & & \\ & & & & \\ R_{2t-1} & & & & \end{bmatrix} \begin{bmatrix} 1 \\ \sigma_2 \\ \sigma_4 \\ \cdot \\ \cdot \\ \cdot \\ \\ \end{bmatrix} \tag{8}$$

The upper right corner element in the matrix of Eqs. (7) is R_1 if t is even or R_3 if t is odd.

Once Eqs. (7) are solved for the even numbered σ_i, then Eqs. (8) give the odd numbered σ_i directly. Thus we have reduced the problem of solving Eqs. (1) to the problem of solving Eqs. (7). The number of galois field operations involved in the transformation (Eqs. 2) and in completing the solution (Eqs. 8) grows only as the square of t, while the number of operations involved in the solution of Eqs. (1) or (7) is proportional to the cube of the number of equations. By reducing this number by a factor of two, we reduce the number of operations by a factor of eight. The extra storage required by Eqs. (2) and (8) is linear in t, while the storage space required to solve Eqs. (1) or (7) is quadratic in t, so our transformation reduces the storage requirements by a factor of four.

The determinant of the $[t/2] \times [t/2]$ submatrix consisting of the last $[t/2]$ columns of the matrix in Eqs. (7) is denoted by N_t. Notice that $N_t = M_t$. Thus, decoding schemes based on repeated evaluations of M_t can profitably employ our transformation.

The steps of our transformation may be summarized as follows:

Given $S_1, S_3, \cdots S_{2t-1}$:

(1) Compute $S_2, S_4, \cdots S_{2t}$ from $S_{2i} = S_i^2$;

(2) Compute $A_1, A_2, \cdots A_{[t/2]}$ from Eq. (6);

(3) Compute $R_1, R_3 \cdots R_{2t-1}$ from Eq. (2).

Thanks to the following lemma, steps (2) and (3) require only half as much work as it might appear.

Lemma. $A_{2I} = 0$ for all $I \neq 0$.

Proof. From (6), for $i > 0$,

$$A_I = \sum_{k=0}^{i-1} A_i S_{i-k}^2 = \left[\sum_{k=0}^{i-1} (A_k)^{1/2} S_{i-k}\right]^2. \tag{9}$$

$$\text{Set } B_i = (A_i)^{1/2}; A_i = B_i^2.$$

Then B_i is defined recursively by

$$B_{-i} = 0 \text{ for } i > 0;$$

$$B_0 = 1;$$

$$B_i = \sum_{k=0}^{i-1} B_k S_{i-k} = S_i + \sum_{k=1}^{i-1} B_k S_{i-k}. \tag{10}$$

Substituting Eq. (10) into itself gives

$$B_i = S_i + \sum_{k=1}^{i-1}\left(\sum_{j=0}^{k-1} B_j S_{k-j}\right) S_{i-k} ,$$

$$B_i = S_i + \sum_{0 \le j < k \le i-1} B_j S_{k-j} S_{i-k} , \tag{11}$$

$$B_i = S_i + \sum_{j=0}^{i-2} B_j \sum_{k=j+1}^{i-1} S_{k-j} S_{i-k} .$$

But

$$\sum_{k=j+1}^{i-1} S_{k-j} S_{i-k} = \begin{cases} 0 \text{ if } i \not\equiv j \bmod 2, \\ S^2_{\left(\frac{i-j}{2}\right)} \text{ if } i \equiv j \bmod 2 . \end{cases} \tag{12}$$

So

$$B_i = S_i + \sum_{j \equiv i \bmod 2}^{i-2} B_j S^2_{\left(\frac{i-j}{2}\right)},$$

$$B_{2I} = S_{2I} + \sum_{J=0}^{I-1} B_{2J} S^2_{I-J} . \tag{13}$$

Iterating gives

$$B_0 = 1$$

$$B_2 = S_2 + S_1^2 = 0 ,$$

$$B_4 = S_4 + S_2^2 + 0 \cdot S_1^2 = 0 .$$

By induction, if $B_{2K} = 0$ for all $0<K<I$, then $B_{2I} = S_{2I} + S_I^2 = 0$, which completes the proof.

The computation of the odd-numbered A_i might be based on Eqs. (9) and (13) instead of Eq. (6). For small t, one has

$$\begin{aligned} B_0 &= 1 \\ B_1 &= S_1 \\ B_3 &= S_3 + B_1 S_2 = S_3 + S_1 S_2 \\ B_5 &= S_5 + B_1 S_4 + B_3 S_2 = S_5 + S_2 S_3 \\ &\vdots \end{aligned} \tag{14}$$

$$\begin{aligned} R_0 &= 1 \\ R_1 &= S_1 \\ R_3 &= S_3 + S_1^2 S_1 \\ R_5 &= S_5 + S_1^2 S_3 \\ R_7 &= S_7 + S_1^2 S_5 + (S_3^2 + S_1^2 S_2^2) S_1 \\ R_9 &= S_9 + S_1^2 S_7 + (S_3^2 + S_1^2 S_2^2) S_3 \\ &\vdots \end{aligned} \tag{15}$$

Since Eqs. (7) and (8) are equivalent to Newton's identities [Eqs. (1)], there is no need to retain the S's and the A's once the R's have been computed.

If the S's were desired for some other purpose, they could be recovered from the R's. For from Eq. (2),

$$S_{2i+1} = R_{2i+1} + \sum_{k=1}^{2i} A_k S_{2i+1-k} , \tag{16}$$

which gives the S's in terms of the R's.

Eq. (6) gives $A_1, \cdots A_{2i-1}$ in terms of $S_1^2, S_2^2, \cdots S_{2i-2}^2$; so Eqs. (16) and (6) define the S's recursively in terms of the R's. This shows that our transformation is invertible.

In addition to the large computational advantage, our transformation also has certain theoretical advantages, arising from the fact that the matrix in Eqs. (7) has constant diagonals. Such matrices (called *persymmetric* matrices) have been enumerated by rank by Daykin (Ref. 10) and Berlekamp (Ref. 11), and from these results it is possible to reach certain conclusions about the distribution of syndromes corresponding to correctable error patterns having more than t errors. A few preliminary results along these lines are given by Berlekamp (Ref. 12).

It is not yet known whether the constant diagonal property may be exploited to reduce even further the computation required to solve Eqs. (7) or to evaluate the determinant N_t. A somewhat similar problem for Toeplitz matrices over the complex field has recently been solved by Trench (Ref. 13), via an algorithm which requires only Kt^2 multiplications to invert a $t \times t$ Toeplitz matrix.

To deal with Newton's identities over a field of characteristic p, one can still convolve the S_k with a sequence A_k to obtain another sequence R_k. In this case the sequence A_k is chosen so that $R_{pi} = 0$ for all $i \neq 0$. The σ_j

for which $f \equiv 1 \bmod p$ can then be expressed directly in terms of the other α's via equations similar to Eqs. (8) above. To obtain the rest of the α's, one must solve a system of Eqs. (7) above. The order of this system is $[(p-1)\,t/p]$. For large p, this reduction is much less substantial than in the binary case.

This decrease in decoder complexity is not directly applicable to the teletype coding project being done by Section 331 (see "Teletype Coding Experiment" in *SPS 37-31*, Vol. III, and in preceding issues, Vol. III). The problem is that the codes to be used for teletype are not over the 2 element field but rather over the 16 element field. A generalization of the result presented here to fields of order 2^r, $r > 1$, would thus be very useful.

D. Estimating the Proportions in a Mixture of Two Normal Distributions using Quantiles

I. Eisenberger[2]

If a random variable is such that with probability p it comes from a normal distribution with density $N(\mu_1, \sigma_1)$ and with probability $1-p$ comes from a normal distribution with density $N(\mu_2, \sigma_2)$, then the distribution of the parent population is said to be *a mixture* of two normal distributions with proportions p and $1-p$. For example, particles may be coming from two sources with different energy spectra. This article is concerned with the problem of estimating p using a small number of sample quantiles when the parameters of the normal distributions are known and the sample size is large. The system is usable with a quantiler for on-board data compression in deep space probes (Ref. 14).

As is typical with the use of sample quantiles for estimation and prediction (Ref. 15–17), the variances of the estimators of p that will be given depend upon the order of the quantiles chosen. An additional complication, however, is that for this type of quantile estimator the variances also depend upon the unknown value of p; and thus one cannot specify *a priori* which quantiles will give optimum results for given values of the parameters of the two normal distributions. Nevertheless, an investigation of six cases involving six different sets of values μ_1, σ_1, μ_2, σ_2 and p (Table 2) indicates that, despite this handicap, estimating p using sample quantiles is feasible. The remainder of this article is concerned with the details and results of this investigation.

Table 2. Parameters of a mixture of two normal distributions

Case	μ_1	σ_1	μ_2	σ_2	p
1	0	1	1	2	0.2
2	0	1	4	3	0.3
3	0	1	5	1	0.05
4	0	1	8	0.5	0.4
5	0	1	0.5	4	0.35
6	0	1	0.5	4	0.015

1. *Review of Quantiles*

To define a *quantile*, consider n independent sample values $x_1, x_2, \cdots, x_n$ taken from a distribution of a continuous type with distribution function $H(x)$ and density function $h(x)$. The s^{th} quantile, or the quantile of order s of the distribution or population, denoted by ζ_s, is defined as the root of the equation $G(\zeta) = s$; that is,

$$s = \int_{-\infty}^{\zeta_s} d\,H(x) = \int_{-\infty}^{\zeta} h(x)\,dx.$$

The corresponding *sample quantile* z_s is defined as follows: If the sample values are arranged in nondecreasing order of magnitude

$$x_{(1)} \lesseqgtr x_{(2)} \lesseqgtr \cdots \lesseqgtr x_{(n)}$$

then $x_{(i)}$ is called the i^{th} *order statistic* and

$$z_s = x_{([ns]+1)}$$

where $[t]$ denotes the greatest integer $\leqq t$.

If $h(x)$ is differentiable in some neighborhood of each quantile value considered, it has been shown (Ref. 18) that the joint distribution of any number of quantiles is asymptotically normal as $n \to \infty$ and that, asymptotically

$$E(z_s) = \zeta_s$$

$$\mathrm{Var}\,(z_s) = \frac{s(1-s)}{nh^2(\zeta_s)}$$

$$\rho_{12} = \left[\frac{s_1(1-s_2)}{s_2(1-s_1)}\right]^{1/2}$$

where ρ_{12} is the correlation between z_{s_1} and z_{s_2}, $s_1 < s_2$.

[2]The author acknowledges the assistance of Thomas Wales, Section 317, in doing some of the computations.

Throughout the remainder of this article we will denote by $F(x)$ and $f(x) = F'(x)$ the distribution and density function, respectively, of the standard normal distribution; that is,

$$F(x) = \int_{-\infty}^{x} f(t)dt\,,\ \text{where}\ f(x) = \frac{1}{(2\pi)^{1/2}}\,e^{-1/2\,x^2}$$

Thus, the density function $g(x)$ of a mixture of two normal distributions is given by

$$g(x) = \frac{p}{\sigma_1(2\pi)^{1/2}}\,exp\left[-\tfrac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right]$$

$$+\frac{1-p}{\sigma_2(2\pi)^{1/2}}\,exp\left[-\tfrac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right]$$

$$=\frac{p}{\sigma_1}f\left(\frac{x-\mu_1}{\sigma_1}\right)+\frac{1-p}{\sigma_2}f\left(\frac{x-\mu_2}{\sigma_2}\right), \qquad (1)$$

and the population quantile ζ of order s is defined by

$$s=\frac{p}{\sigma_1\sqrt{2\pi}}\int_{-\infty}^{\zeta} exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2\right]dx$$

$$+\frac{1-p}{\sigma_2(2\pi)^{1/2}}\int_{-\infty}^{\zeta} exp\left[-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2\right]dx$$

$$=pF\left(\frac{\zeta-\mu_1}{\sigma_1}\right)+(1-p)\,F\left(\frac{\zeta-\mu_2}{\sigma_2}\right). \qquad (2)$$

Since we are assuming a large sample size, the asymptotic distribution of the sample quantiles will also be assumed.

2. The Maximum-Likelihood Estimate of p

In order to compare the results obtained by estimating p using quantiles with the best estimate using all the sample values, the maximum-likelihood estimate of p will be derived, along with the asymptotic distribution of this estimate. The maximum-likelihood estimate of p is obtained by setting equal to zero the derivative with respect to p of the logarithm of the likelihood function and solving the resulting equation for p. Accordingly, one has

$$L(x_1,x_2,\cdots,x_n;p)=\prod_{i=1}^{n}\left[\frac{p}{\sigma_1}f\left(\frac{x_i-\mu_1}{\sigma_1}\right)+\frac{(1-p)}{\sigma_2}f\left(\frac{x_i-\mu_2}{\sigma_2}\right)\right]$$

$$\ln L=\sum_{i=1}^{n}\ln\left[\frac{p}{\sigma_1}f\left(\frac{x_i-\mu_1}{\sigma_1}\right)+\frac{(1-p)}{\sigma_2}f\left(\frac{x_i-\mu_2}{\sigma_2}\right)\right] \qquad (3)$$

$$\frac{\partial \ln L}{\partial p}=\sum_{i=1}^{n}\frac{f\left(\frac{x_i-\mu_1}{\sigma_1}\right)-f\left(\frac{x_i-\mu_2}{\sigma_2}\right)}{\frac{p}{\sigma_1}f\left(\frac{x_i-\mu_1}{\sigma_1}\right)+\frac{(1-p)}{\sigma_2}f\left(\frac{x_i-\mu_2}{\sigma_2}\right)}=0.$$

The maximum-likelihood estimate of p is the value of p which satisfies (3) if this value is between zero and one. If the value is negative, the estimate is zero; and, if the value is greater than one, the estimate is one. Now, neglecting these end effects, maximum-likelihood theory informs us that, for large samples, the distribution of the maximum-likelihood estimate of p is approximately normal with mean equal to p and variance given by

$$\mathrm{Var}\,(p_L)=\frac{-E\left[\frac{\partial^2}{\partial p^2}\ln g\,(x)\right]^{-1}}{n}$$

$$=\frac{1}{n}E\left[\frac{f\left(\frac{x-\mu_1}{\sigma_1}\right)-f\left(\frac{x-\mu_2}{\sigma_2}\right)}{\frac{p}{\sigma_1}f\left(\frac{x-\mu_1}{\sigma_1}\right)+\frac{(1-p)}{\sigma_2}f\left(\frac{x-\mu_2}{\sigma_2}\right)}\right]^{-2}$$

$$=\frac{1}{n}\left\{\int_{-\infty}^{\infty}\frac{\left[f\left(\frac{x-\mu_1}{\sigma_1}\right)-f\left(\frac{x-\mu_2}{\sigma_2}\right)\right]^2dx}{\frac{p}{\sigma_1}f\left(\frac{x-\mu_1}{\sigma_1}\right)+\frac{(1-p)}{\sigma_2}f\left(\frac{x-\mu_2}{\sigma_2}\right)}\right\}^{-1}. \qquad (4)$$

3. Estimating p Using Quantiles

In (2), for a fixed value of s, ζ is defined uniquely. This relationship provides a simple estimator for p using one quantile, which is obtained as follows. Replace ζ in (2) by the corresponding sample quantile z of order s and solve for p. Thus, one obtains

$$\hat{p}_1 = \frac{s - F\left(\dfrac{z - \mu_2}{\sigma_2}\right)}{F\left(\dfrac{z - \mu_1}{\sigma_1}\right) - F\left(\dfrac{z - \mu_2}{\sigma_2}\right)} \tag{5}$$

which is easily computed from a table of the standard normal probability function. An estimator using m quantiles can be constructed as a linear combination of individual estimators as follows

$$\hat{p} = \sum_{i=1}^{m} a_i \hat{p}_i = \sum_{i=1}^{m} \frac{a_i \left[s_i - F\left(\dfrac{z_i - \mu_2}{\sigma_2}\right)\right]}{F\left(\dfrac{z_i - \mu_1}{\sigma_1}\right) - F\left(\dfrac{z_i - \mu_2}{\sigma_2}\right)} \tag{6}$$

where $\sum_{i=1}^{m} a_i = 1$. As in the case of the maximum likelihood estimation, if the computed value of $\hat{p}_1$ from (5) or $\hat{p}$ from (6) is either negative or greater than one, $\hat{p}$ is taken to be zero or one, respectively.

Since the distribution of $\hat{p}$ in either (5) or (6) is extremely difficult to determine, let us tentatively make the following two assumptions, again neglecting end effects.

(1) A good approximation for $E(\hat{p})$ and $E(\hat{p}_1)$ can be had by substituting $E(z_i) = \zeta_i$ for z_i in (5) and (6), obtaining $E(\hat{p}) \cong p$ and $E(\hat{p}_1) \cong p$.

(2) A good approximation for $\mathrm{Var}(p_1)$ is given by

$$\mathrm{Var}(\hat{p}_1) = \left(\frac{\partial \hat{p}_1}{\partial \zeta}\right)^2 \mathrm{Var}(z) \tag{7}$$

and a good approximation $\mathrm{Var}(\hat{p})$ is given by

$$\mathrm{Var}(\hat{p}) = \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j \frac{\partial p_i}{\partial \zeta_i} \frac{\partial p_j}{\partial \zeta_j} \sigma_{ij} \tag{8}$$

where

$$\frac{\partial p_i}{\partial \zeta_i} = \frac{\partial \hat{p}_i}{\partial z_i}\bigg|_{z_i = \zeta_i}$$

and σ_{ij} denotes the covariance of z_i and z_j.

As noted previously, if some knowledge of the actual value of p is not available, the orders of the quantiles which minimize the variance of $\hat{p}$ using quantiles cannot be determined. It is of value, nevertheless, to compare the variances of the optimum estimators of p using a small number of quantiles with the variance of the maximum-likelihood estimator. This information provides a measure of the relative reliability of the best possible quantile estimators when compared with the best estimator using all the sample values. A computer study was therefore made, for each of the cases given in Table 2, to determine the order of the single quantile which minimizes the variance of $\hat{p}_1$ as given by (7), and the order of the two quantiles, as well as the values of a and $1 - a$, which minimize the variance of $\hat{p}$ as given by (8) for $m = 2$. For these optimum orders, the variances of the quantile estimators were computed. The variance of the maximum-likelihood estimator as given by (4) was also computed. The results are shown in Table 3. Assuming the validity of the assumption that (5) and (6) are good approximations for the variance of the quantile estimators, it can be seen that in Cases 1 - 4 the variances of the optimum quantile estimators compare very favorably with those of the maximum-likelihood estimators. And even for Cases 5 and 6 the variances of the quantile estimators, although significantly larger than those of the maximum-likelihood estimators, are sufficiently small so that estimates of p based on one and two sample quantiles are still meaningful.

Another purpose of the computer study was to test the two assumptions made with respect to the $E(\hat{p})$ and $\mathrm{Var}(\hat{p})$ of the quantile estimators. For each case in Table 2, 100 sets of samples each containing 500 sample values were generated from the corresponding distribution. For each set the optimum sample quantiles were chosen for the one quantile and two quantile estimators. Then $\hat{p}_L$, $\hat{p}_1$ and $\hat{p}$ were computed. Thus, for each case this procedure gave 100 sample values of $\hat{p}_L$, $\hat{p}_1$ and $\hat{p}$. The sample means, denoted by $\bar{p}_L$, $\bar{p}_1$ and $\bar{p}$, respectively, and the sample variances were computed for each set of these 100 sample values and are given in Table 4.

By comparing $\bar{p}_1$ and $\bar{p}$ with the actual value of p in each case, it can be seen that the assumption that the quantile estimators of p are approximately unbiased cannot be rejected. In addition, by comparing the sample variances of the quantile estimators in Table 4 with the theoretical variances of the estimators as computed from (7) and (8) and given in Table 3, it is also evident that these expressions provide good approximations for $\mathrm{Var}(\hat{p}_1)$ and $\mathrm{Var}(\hat{p})$, respectively.

Table 3. Optimum orders for estimators of p using one and two quantiles and minimum variances

$n = 500$

Case	One quantile		Two quantile				M.L.
	Opt. s	Var $(\hat{p}_1)$	Opt. s_1	Opt. s_2	Opt. α	Var $(\hat{p})$	Var $(\hat{p}_L)$
1	0.6411	0.003949	0.5692	0.7178	0.4961	0.003643	0.002705
2	0.4281	0.0009122	0.3811	0.4706	0.5026	0.0008726	0.0008110
3	0.0496	0.00009991	0.0482	0.0511	0.4989	0.00009910	0.00009877
4	0.4000	0.0004800[a]	0.3800[a]	0.4200[a]	0.4154[a]	0.0004761[a]	0.0004800[a]
5	0.7613	0.003308	0.7358	0.7884	0.5055	0.003193	0.001291
6	0.6344	0.004101	0.6144	0.6700	0.5499	0.003985	[illegible]

[a] Due to numerical difficulties, the optimality of these results is uncertain.

Table 4. Sample means and variances of the maximum-likelihood estimator and quantile estimators of p, for 100 sample values

$n = 500$

Case	p	$\bar{p}_L$	$\bar{p}_1$	$\bar{p}$	Sample Var $(\hat{p})$		
					M.L.	1 quant.	2 quant
1	0.20	0.1993	0.2058	0.2116	0.002498	0.004040	0.003602
2	0.30	0.3017	0.2992	0.3017	0.0007097	0.0009193	0.0008673
3	0.05	0.0500	0.0508	0.0496	0.0001019	0.0001411	0.0001364
4	0.40	0.4000	0.4005	0.4016	0.0003712	0.0003637	0.0004781
5	0.35	0.3468	0.3555	0.3598	0.001114	0.003032	0.002972
6	0.015	0.0132	0.0198	0.0221	0.0007945	0.003698	0.003776

The end effects mentioned previously, and which have been neglected in our analyses, are negligible unless p is very close to zero or one, as in Case 6. Of the six cases considered, only in Case 6 were any of the estimates of p negative. In fact, about a third of each set of 100 estimates $\hat{p}_L$, $\hat{p}_1$ and $\hat{p}$ were negative for this case. Since the end effects were not taken into account in the theoretical expressions for the expected values and variances of the estimators, the negative estimates were not corrected to zero in computing the sample means and variances. Corrections for end effects, when they are significant, would result in a larger sample mean for p close to zero, a smaller sample mean when p is close to one, and smaller sample variances. To this extent, then, the theoretical expected values of the estimators, as given previously, are biased; and the approximations to the variances, as given by (4), (7) and (8), are not entirely valid. Corrections to take into account these end effects can be made but will be omitted.

Although it should be evident from our theoretical and empirical results that *optimum* one and two quantile estimators of p give meaningful results when compared with maximum-likelihood estimators, the question still remains as to what procedure to adopt when p is entirely unknown and one wishes to estimate p optimally, in some sense, using quantiles. Unfortunately, in the one and two quantile cases, a poor choice of quantiles can have a drastic effect on the variance of the estimators. For example, in Case 1 if one uses the sample quantile of order 0.2 instead of the optimum one of order 0.6411, Var $(\hat{p})$ increases by a factor of about 58, from 0.00395 to 0.2282. In the two quantile case, since neither s_1 nor s_2 differs much from s and they are approximately symmetric about s, a poor choice can also increase the variance of the estimator by a large factor. Moreover, increasing the number of quantiles does not automatically guarantee better results. A poor choice of four, six or even more quantiles can result in a larger variance than would be obtained from the use of a single optimum quantile. Thus, if the primary goal is the determination of the orders of the m quantiles which will provide the "best" estimates according to some reasonable criterion such as minimizing the variance when p is entirely unknown, a methodical investigation of a function with five parameters is essential. And this can indeed be a formidable task.

If, on the other hand, one considers this problem from the viewpoint of data compression of space telemetry, one sees that it is intimately related to the overall problem of using sample quantiles effectively (but no longer necessarily optimally) for statistical estimation and prediction in connection with space experiments. In this context, one possible solution of the problem considered here is already at hand. It remains to be seen whether this proposed solution is a satisfactory one.

The investigation into the use of sample quantiles in statistical procedures was initiated because of a desire to achieve a significant amount of data compression of space telemetry. Optimum unbiased estimators of the mean μ and standard deviation σ of a normal distribution were first constructed using up to 10 pairs of symmetric quantiles. These are described in Ref. 15. Since for the use of this system in space experiments the order of the quantiles would be specified in advance, and the optimum quantiles for estimating μ are not optimum for estimating σ, a compromise had to be made. The orders of the symmetric quantiles which minimized the linear combination $\mathrm{Var}(\hat{\mu}) + b\,\mathrm{Var}(\hat{\sigma})$ were determined for $b = 1, 2, 3$ and are designated as suboptimum quantiles. Unbiased estimators of both μ and σ using these suboptimum quantiles were then constructed. The same compromise was used in the various tests of hypotheses described in Ref. 15–17 for $k = 4$, (k denotes the number of quantiles used). This compromise is proposed for estimating p, and the discussion will be restricted to $k = 4$, $b = 1$.

For $b = 1$ and $k = 4$, the orders of the suboptimum quantiles are as follows:

$$s_1 = 0.0668 \qquad s_2 = 0.2912$$

$$s_3 = 0.7088 \qquad s_4 = 0.9332$$

Using four quantiles the estimator of p is given by

$$\tilde{p} = \frac{1}{4}\sum_{i=1}^{4} \hat{p}_i, \text{ where}$$

$$\hat{p}_i = \frac{s_i - F\left(\dfrac{z_i - \mu_2}{\sigma_2}\right)}{F\left(\dfrac{z_i - \mu_1}{\sigma_1}\right) - F\left(\dfrac{z_i - \mu_2}{\sigma_2}\right)}, \quad i = 1, 2, 3, 4$$

Again neglecting end effects, the $\mathrm{Var}(\tilde{p})$ is approximated by

$$\mathrm{Var}(\tilde{p}) \cong \frac{1}{16}\sum_{i=1}^{4}\sum_{j=1}^{4} \frac{\partial \hat{p}_i}{\partial \zeta_i}\frac{\partial \hat{p}_j}{\partial \zeta_j}\rho_{ij}\left[\mathrm{Var}(z_i)\right]^{1/2}\left[\mathrm{Var}(z_j)\right]^{1/2} \tag{9}$$

where

$$\frac{\partial \hat{p}}{\partial \zeta_i} = \frac{-\dfrac{1}{\sigma_2} f\left(\dfrac{\zeta_i - \mu_2}{\sigma_2}\right)\left[F\left(\dfrac{\zeta_i - \mu_1}{\sigma_1}\right) - F\left(\dfrac{\zeta_i - \mu_2}{\sigma_2}\right)\right] - \left[s_i - F\left(\dfrac{\zeta_i - \mu_2}{\sigma_2}\right)\right]\left[\dfrac{1}{\sigma_1} f\left(\dfrac{\zeta_i - \mu_1}{\sigma_1}\right) - \dfrac{1}{\sigma_2} f\left(\dfrac{\zeta_i - \mu_2}{\sigma_2}\right)\right]}{\left[F\left(\dfrac{\zeta_i - \mu_1}{\sigma_1}\right) - F\left(\dfrac{\zeta_i - \mu_2}{\sigma_2}\right)\right]^2}$$

$$\mathrm{Var}(z_i) = \frac{s_i(1 - s_i)}{n\left[\dfrac{p}{\sigma_1} f\left(\dfrac{\zeta_i - \mu_1}{\sigma_1}\right) + \dfrac{1 - p}{\sigma_2} f\left(\dfrac{\zeta_i - \mu_2}{\sigma_2}\right)\right]^2}$$

and

$$\rho_{ij} = \left[\frac{s_i(1 - s_j)}{s_j(1 - s_i)}\right]^{1/2}, \quad i < j$$

Table 5. Comparison of Var(p) and Var′(p)

$n = 500$

Case	Var ($\tilde{p}$)	Var ($\hat{p}$)	$\frac{\text{Var}(\tilde{p})}{\text{Var}(\hat{p})}$
1	0.006979	0.003643	1.92
2	0.003321	0.0008726	3.81
3	0.002651	0.00009919	26.74
4	0.001773	0.0004761	3.72
5	0.002551	0.003193	0.80
6	0.003192	0.003985	0.80

Using (9) the variances of the estimators using suboptimum quantiles were computed for the six cases and are given in Table 5. Var($\hat{p}$) is also shown again for comparison.

As might be expected, the results for the six cases vary. However, only for Case 3 is the result quite poor, quantitatively speaking, and this is due to the fact that opt. $s = 0.0496$ is smaller than all four of the orders of the suboptimum quantiles. Nevertheless, even for this case, since the standard deviation of $\tilde{p}$ is 0.0515, there is considerable probability that a given p will at least provide a basis for the qualitative conclusion that p is small. It should also be noted that Var($\tilde{p}$), in common with Var($\hat{p}$), Var($\hat{p}_1$) and Var($\hat{p}_2$), is inversely proportional to n. Thus, for sufficiently large n, even in Case 3 the suboptimum estimator of p will also be quantitatively significant. And large sample sizes are not uncommon in space experiments.

E. Instantaneously Synchronizable Block Code Dictionaries

J. J. Stiffler

1. Introduction

Let D(N) be a block code dictionary consisting of N-symbol words defined over an r-symbol alphabet. The code is said to be instantaneously synchronizable if the knowledge of any consecutive $2N$-1 symbols in a sequence of code words is sufficient to determine synchronization (i.e., to establish which symbol of each word is the initial symbol). Comma-free codes (Ref. 19) are instantaneously synchronizable as well as prefix codes (Ref. 20). These and other constructions will be discussed, and the number of words in the dictionaries for a given N will be compared for the various techniques.

2. Several Constructions

Comma-free codes have the property that no N-tuple consisting of the last k symbols of one code word followed by the first N-k symbols of another (not necessarily different) word can be a code word for any k, $1 \leq k \leq N-1$. Such code dictionaries are clearly instantaneously synchronizable. They have two disadvantages, however. First, it is necessary to observe the symbols in groups of N and to determine if they are code words or not. This can be time consuming if N is large. Second, the encoding and decoding procedures for comma-free codes are generally quite complex.

The first of these disadvantages is effectively nonexistent in a more restricted class of codes: the prefix codes. All words in a prefix code dictionary begin with the same m-symbol prefix and are so constrained that, regardless of the word sequence, this prefix can occur *only* at the beginning of a code word. Thus, when the m-symbol prefix is observed it may immediately be concluded that the initial code word symbols have been found.

Another construction which results in an even more restricted class of codes, but one for which both of the above objectives to comma-free codes are to some extent overcome, is the following: An m-symbol prefix is reserved as a marker or "comma" as before, but now the constraint length is shortened from N to some smaller integer n. The word length N is taken to be the number of symbols separating the initial symbols of two successive occurrences of the comma. Thus, if the comma is transmitted after every k^{th} n-tuple, $N = m + kn$. The n-tuples must be so constrained that no m-tuple formed from any concatenation of allowable n-tuples or from any n-tuple and the comma itself can be the comma. Since the n-tuples can be selected independently, the encoding and decoding processes can be kept within reasonable bounds even for large values of N.

Some special cases of this last construction are: (1) $m = 1$; one symbol must be set aside for a comma and not used in any of the code words. There are $(r-1)^n$ possible n-tuples which may be used. The encoding is extremely simple in this case. (2) $m = n$; one n-tuple is used as a comma. The number of n-tuples which may be used as code words has been determined in Ref. 21. When n is even it is possible to use

$$r^n\left(1 - \frac{r}{r-1}\frac{1}{r^{n/2}} + \frac{1}{(r-1)r^n}\right) \tag{1}$$

such n-tuples, and when n is odd, the number becomes

$$r^n\left(1 - \frac{2r-1}{r-1}\frac{1}{r^{\frac{n+1}{2}}} + \frac{1}{(r-1)r^n}\right) \tag{2}$$

(3) $m = 2n$; two consecutive n-tuples are used to define the comma. If the two-n-tuples are taken to be

$$a\,a \cdots\cdots\cdots\, a\,a$$

and

$$a\,a \cdots\cdots\cdots\, a\,b$$

respectively, where a and b are any two of the r symbols of the alphabet, it is evident that only one n-tuple need be excluded from the set used to transmit data; viz the n-tuple $aa \cdots\cdots a$. This is apparent when it is observed that if the $2n$-tuple

$$\underbrace{a\,a \cdots\cdots a}_{n\text{-symbols}}\;\underbrace{a\,a \cdots\cdots a\,b}_{n\text{-symbols}}$$

is to be found within the data sequence, then the last ℓ symbols of one n-tuple, followed by a second n-tuple which is in turn followed by the first n-ℓ symbols of a third n-tuple must result in this sequence of symbols. But either $1 \leqslant \ell \leqslant n-1$, in which case the middle n-tuple must be the prohibited $aa \cdots\cdots a$, or else $\ell = 0$, and this n-tuple must occur initially. Hence, $r^n - 1$ n-tuples can be used to form the data portion of the code words. The encoding and decoding procedures here are nearly as simple as those encountered when a single symbol is used as a comma. It is only necessary to exclude the n-tuple $aa \cdots\cdots a$ from the data sequence (and then only when it occurs as a subword, not as an overlap of two such subwords). When this is done, the n-tuples may be transmitted directly.

3. Comparison of Dictionary Sizes for Large N

All instantaneously synchronizable codes clearly must have the property that if an N-tuple is in the dictionary then none of its cyclic permutations can be in the dictionary. If an N-tuple has a periodicity less than N it is a cyclic permutation of itself and cannot be in the dictionary. Thus no instantaneously synchronizable dictionary can have more than

$$W(N,r) = \frac{1}{N}\sum_{d/n}\mu(d)\; r^{n/d} \tag{3}$$

words where $\mu(d)$ is the Mobius function defined as follows: Let $d = p_1^{\alpha_1}p_2^{\alpha_2}\cdots p_t^{\alpha_t}$ be the factorization of d into prime powers. Then

$$\mu(d) = \begin{cases} 1 & \text{if } d = 1 \\ (-1)^t & \text{if } \alpha_1 = \alpha_2 = \cdots = \alpha_t = 1 \\ 0 & \text{if } \alpha_i > 1 \text{ for any } i \end{cases} \tag{4}$$

(cf Ref. 19). It is known that comma-free codes can be constructed achieving this bound for odd lengths N. When N is even comma-free codes containing at least $2/e\; r^N/N$ words can be constructed (Ref. 22).

It is shown in Ref. 20 that the number of words $P(N, r)$ in a prefix dictionary is for $r = 2$ and asymptotically with N,

$$\frac{1}{e}\frac{2^N}{N} \geqslant P(N,2) \geqslant \frac{1}{2\log_2 e}\cdot\frac{2^N}{N}. \tag{5}$$

This result can be generalized, for arbitrary r, to establish that

$$\frac{1}{e}\frac{r^N}{N} \geqslant P(N,r) \geqslant \frac{1}{(r-1)\,r^{\frac{1}{r-1}}\log_r e}\,\frac{r^N}{N}. \tag{6}$$

Asymptotic bounds on the number of words $C(N,r)$ in large comma dictionaries are also readily obtained. When $m = 1$, the number of words is $C_1(N,r) = (r-1)^{N-1}$ regardless of the value of n. For purposes of comparison this may be written

$$C_1(N,r) = \frac{N\left(1 - \frac{1}{r}\right)^N}{r-1}\,\frac{r^N}{N}. \tag{7}$$

When $m = n$ we have, for even n,

$$C_n(N,r) = \left[r^n\left(1 - \frac{r}{r-1}\frac{1}{r^{n/2}} + \frac{1}{(r-1)r^n}\right)\right]^{\frac{N}{n}-1} \tag{8}$$

since $N = (k+1)n$. Thus,

$$C_n(N,r) = \frac{r^N}{r^n}\left[1 - \frac{r}{r-1}\frac{1}{r^{n/2}}\left(1 - \frac{1}{r^{n/2+1}}\right)\right]^{\frac{N}{n}-1}. \tag{9}$$

We define a new function of N, $a = a(N)$ by the relationship

$$n\,r^{n/2} = N/a. \tag{10}$$

Then,

$$C_n(N, r) = \frac{r^N}{r^n}\left[1 - \frac{r}{r-1}\frac{an}{N}\left(1 - \frac{an}{rN}\right)\right]^{N/n-1} \quad (11)$$

which for large N yields

$$C_n(N, r) \sim \frac{r^N}{r^n} exp\left\{-\frac{r}{r-1}a\right\} =$$

$$r^N exp\left\{-\left(n \log_e r + \frac{r}{r-1}\frac{N}{nr^{n/2}}\right)\right\}. \quad (12)$$

It is easily verified that

$$exp\left\{-\left(n \log_e r + \frac{r}{r-1}\frac{N}{nr^{r/2}}\right)\right\}$$

monotonically increases with n for

$$\log_e r < \frac{r}{r-1}\frac{N}{nr^{n/2}}\left(\frac{1}{n} + \frac{1}{2}\log_e r\right) \quad (13)$$

and is a monotonically decreasing function of n otherwise. Hence, a maximum occurs when

$$\log_e r = \frac{r}{r-1}a\left(\frac{1}{n} + \frac{1}{2}\log_e r\right) \sim \frac{r}{r-1}\frac{a}{2}\log_e r. \quad (14)$$

Thus, for the dictionary size to be maximum

$$a = \frac{N}{nr^{n/2}} \approx \frac{2(r-1)}{r}$$

and, for large n, is essentially constant. To maximize the size of the dictionary, then, for a given constraint length n it is required that

$$N = 2(r-1)n\, r^{(n/2-1)}. \quad (15)$$

Since

$$r^{n/2} < nr^{n/2} < r^{n/2(1+\epsilon)} \quad (16)$$

for sufficiently large n, and for any $\epsilon > 0$, it follows that, asymptotically, when N satisfies Eq. (15),

$$\left(\frac{rN}{2(r-1)}\right)^{\frac{2}{1+\epsilon}} < r^n < \left(\frac{rN}{2(r-1)}\right)^2 \quad (17)$$

and from Eq. (12),

$$C_n\left[N = 2(r-1)\, nr^{(n/2-1)}, r\right] \sim \left[\frac{2(r-1)}{er}\right]^2 \frac{r^N}{N^2}. \quad (18)$$

When n is odd a similar argument establishes that

$$C_n\left[N = \frac{2(r-1)}{2r-1} nr^{\frac{n+1}{2}}, r\right] \sim r\left[\frac{2(r-1)}{e(2r-1)}\right]^2 \frac{r^N}{N^2}. \quad (19)$$

Finally, when $m = 2n$,

$$C_{2n}(N,r) = (r^n - 1)^{\frac{N}{n}-2}$$

$$= \frac{r^N}{r^{2n}}\left(1 - \frac{1}{r^n}\right)^{\frac{n}{N}-2}. \quad (20)$$

Defining

$$nr^n = \frac{N}{a}, \quad (21)$$

we find that

$$C_{2n}(N,r) \sim \frac{r^N}{r^{2n}} e^{-a} = r^N exp\left[-\left(\frac{N}{nr^n} + 2n \log_e r\right)\right] \quad (22)$$

The expression $N/(nr)^n + 2n \log_e r$ is minimum when $a = 2$. Further, since for sufficiently large n,

$$r^n < \frac{N}{2} < r^{n(1+\epsilon)} \quad (23)$$

when $N = 2nr^n$, and for any $\epsilon > 0$, it follows that

$$\left(\frac{N}{2}\right)^{\frac{1}{1+\epsilon}} < r^n < \frac{N}{2} \quad (24)$$

and

$$C_{2n}(N = 2nr^n, r) \sim \frac{4}{e^2}\frac{r^N}{N^2}. \quad (25)$$

It should be emphasized that these asymptotic results [Eqs. (18), (19) and (25)] are not valid for every value of N, but rather only when N is the proper multiple of nr^n. Thus, for arbitrarily large values of N, the code dictionary may be much smaller than r^N/N^2. Eq. (25) holds for large integer values of n and Eq. (18) for large even integers n. Eq. (19), on the other hand, is valid only for certain odd values of n, since

$$N = \frac{2(r-1)}{2r-1} nr^{\frac{n+1}{2}}$$

must also be an integer.

In conclusion, the comma codes discussed here are always less efficient than prefix codes and comma-free codes, the former having, asymptotically with the word length N, only $1/N^{th}$ as many code words as the latter. Nevertheless, the simplicity of encoding and decoding using these codes may more than make up for the defect.

References

1. Moore, E. F., *Sequential Machines, Selected Papers*, Addison-Wesley, Reading, Mass., 1964.
2. Golomb, S. W., and Welch, L. R., *Non-Linear Shift Register Sequences*, Technical Memorandum No. 20-149, Jet Propulsion Laboratory, Pasadena, California, 1957.
3. Easterling, M., "A Skin-Tracking Radar Experiment Involving the COURIER Satellite," Trans. *IRE, Space Electronics and Telemetry*, Vol. SET-8 1962, Pp. 76–84.
4. Bose, R. C., and Ray-Chaudhuri, D. K., "On a class of Error-correcting Binary Group Codes," *Information and Control 3*, 1960, pp. 68–79.
5. Bose, R. C., and Ray-Chaudhuri, D. K., "Further results on Error-correcting Binary Group Codes," *Information and Control 3*, 1960, pp. 279–290.
6. Chien, R. T., "Cyclic Decoding Procedures for Bose-Chaudhuri-Hocquenghem Codes," *IEEE Profession Group on Information Theory, 1964*, pp. 357–362.
7. Peterson, W. W., *Error-Correcting Codes*, John Wiley and Sons, New York, 1961, p. 176.
8. Massey, J. L., private communication, 1965.
9. Solomon, G., SPS 37-29, Vol. 4, Jet Propulsion Laboratory, Pasadena, 1965, pp. 296–298. See also Fredricksen, H. N. Ibid, p. 299.
10. Forney, D., private communication, 1965.
11. Daykin, G. E., "Distribution of Bordered Persymmetric Matrices in a Finite Field," *Journal für die reine und angewandte Mathematik* 203, 1960, pp. 47–54.
12. Berlekamp, E. R., *The Enumeration of Matrices by Rank*, Bell Telephone Laboratories, Murray Hill, N. J., 1963.
13. Berlekamp, E. R., "The Distribution of Matrices Resulting from Newton's Identities in a Field of Characteristic Two," in *Notes on System Theory, VII*, ERL, University of California, Berkeley, 1965.
14. Trench, W. F., "An Algorithm for the Inversion of Finite Toeplitz Matrices," *Journal of the Society of Industrial and Applied Mathematics*, Vol. 12, 1965, pp. 515–522.
15. *Quantile System*, Space Program Summaries No. 37-32, Vol. 3, Jet Propulsion Laboratory, Pasadena (to be published).
16. Eisenberger, I., and Posner, E. C., *Systematic Statistics Used for Data Compression in Space Telemetry*, Technical Report No. 32-510, Jet Propulsion Laboratory, Pasadena, Oct. 1, 1963.
17. Eisenberger, I., *Tests of Hypotheses and Estimation of the Correlation Coefficient Using Quantiles*, I, Technical Report, Jet Propulsion Laboratory, Pasadena (to be published).
18. Eisenberger, I., *Tests of Hypotheses and Estimation of the Correlation Coefficient Using Quantiles*, II, Technical Report, Jet Propulsion Laboratory, Pasadena (to be published).

References (Cont'd)

19. Cramér, H., *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N. J., 1946, pp. 367–370.

20. Golomb, S. W., Gordon, B., and Welch, L. R., "Comma-Free Codes," *Canadian Journal of Mathematics*, Vol. 10, No. 2, 1958, pp. 202–209.

21. Gilbert, E. N., "Synchronization of Binary Messages," IRE Transactions on Information Theory, IT-6, 1960.

22. Kendall, W. B., "Optimum Synchronizing Words for Fixed Word-Length Code Dictionaries" (to be published).

23. Jiggs, B. H., "Recent Results in Comma-Free Codes," *Canadian Journal of Mathematics*.

XXV. Communications Systems Research: Planetary Radar

A. Solar Radar Occultations

P. Reichley and D. O. Muhleman

In a recent article (Ref. 1, pp. 229–239) we studied the effect of the solar corona on radar observations of Venus near superior conjunction. As the radar beam travels to and from Venus the electron plasma of the corona in conjunction with the relative motion of Venus creates a doppler shift and delay in range measurement. This effect generates data for studies of the size and nature of the corona. Our preliminary conclusion in Ref. 1 was that in the 12½-cm wavelength range, used by the JPL radar astronomy program, this effect is measureable. Full understanding of it is important in precision tracking of space probes in this region of the solar system. This conclusion was reached using a simplified model of the refractive index of the corona. The effect is studied in this article using the accepted model of the corona, based on the measured coronal electron density profile. Circular orbits for both Earth and Venus are assumed throughout.

1. Introduction

The accepted model of the coronal electron density profile leads to the index of refraction of the corona [Ref. 1, Eq. (11), p. 233] at 2388 Mc

$$n^2(r) = 1 - \alpha\left(\frac{\beta}{r^6} + \frac{\gamma}{r^{16}}\right) \tag{1}$$

where

$\alpha = 1.4108 \times 10^{-3}$ = function of frequency and coronal properties.

$\beta = 1.55$

$\gamma = 2.99$

and r is measured in solar radii. The round-trip range difference (relative to the straight-line range) due to refraction is given by [Ref. 1, Eq. (19), p. 234]

$$\delta\rho = 2\left\{\int_{r_m}^{r_\oplus} \frac{r\,dr}{[r^2n^2(r) - b^2]^{1/2}} + \int_{r_m}^{r_♀} \frac{r\,dr}{[r^2n^2(r) - b^2]^{1/2}} - \left(r_\oplus^2 + r_♀^2 - 2r_\oplus r_♀ \cos\theta_d\right)^{1/2}\right\} \tag{2}$$

where

$r_\oplus$ = radius vector to Earth from the Sun.

r_φ = radius vector to Venus from the Sun.

r_m = radius vector to point of closest approach of the ray to the Sun.

θ_d = angular separation of Earth and Venus with respect to the Sun.

b = impact parameter of the ray given by $r_\oplus$, r_φ and θ_d.

The doppler shift due to refraction is given by [Ref. 1, Eq. (29), p. 238] since

$$\dot{\rho} = \dot{\theta}_d b$$

and

$$\dot{\delta\rho} = \dot{\theta}_d (b - b_0)$$

where b_0 is the impact parameter in the absence of a refracting medium. The doppler frequency effect of the refraction is

$$\delta f_r(t) = \frac{-2f_0\omega_s}{c}\left\{b(t) - \frac{r_\oplus r_\varphi \sin(\omega_s t)}{[r_\oplus^2 + r_\varphi^2 - 2r_\oplus r_\varphi \cos(\omega_s t)]^{1/2}}\right\}, \qquad (3)$$

where

f_0 = transmitter frequency

ω_s = synodic angular rate of Venus

c = speed of light

and $b(t)$ and $\omega_s t = \theta_d(t)$ are the impact parameter and angular separation, respectively.

2. Solution for the Impact Parameter

To solve for the impact parameter we use the ray equation [Ref. 1, Eq. (2), p. 231] and the locations of Venus and Earth given by θ_d, $r_\oplus$ and r_φ,

$$\frac{d\theta}{dr} = \frac{b}{r\left[r^2n^2(r) - b^2\right]^{1/2}} \qquad (4)$$

$$\theta(r_\varphi) = \theta_\varphi,\ \theta(r_\oplus) = \theta_\oplus,\ \theta_d = \theta_\varphi + \theta_\oplus$$

We use an iterative scheme to find r_m, the closest approach, which is related to b by Bouger's Theorem,

$$b = r_m n(r_m) \qquad (5)$$

The scheme used is given by

$$(r_m)_{k+1} = \frac{(r_\oplus^2 + r_\varphi^2 - 2r_\varphi r_\oplus \cos\theta_k)^{1/2}}{r_\oplus r_\varphi \sin\theta_k}(r_m)_k (r_m)_0 \qquad (6)$$

where

$$\theta_k = \int_{(r_m)_k}^{r_\oplus} \frac{b_k dr}{r\left[r^2n^2(r) - b_k^2\right]^{1/2}} + \int_{(r_m)_k}^{r_\varphi} \frac{b_k dr}{r\left[r^2n^2(r) - b_k^2\right]^{1/2}}$$

$$b_k = (r_m)_k\, n\left[(r_m)_k\right]$$

$$(r_m)_0 = \frac{r_\varphi r_\oplus \sin\theta_d}{(r_\varphi^2 + r_\oplus^2 - 2r_\oplus r_\varphi \cos\theta_d)^{1/2}}$$

k = iteration number.

For the worst case, i.e., the ray approaching closest to the Sun, 11 iterations were required for 8-place accuracy. The idea put forward in Ref. 1 to solve for b as a function of θ_d by a least squares fit was discarded, due to the double-valuedness of θ_d with respect to b (Fig. 1). The iterative scheme, Eq. (6), is, therefore, good for only that part of the curve in Fig. 1 which corresponds to $r_m \geq 1.286\, r_\odot$.

The double-valuedness of θ_d then shows that the index, Eq. (1), yields two ray paths between the Earth and Venus for $177.81° \leq \theta_d \leq 179.07°$, which is almost the complete range of interest. When $179.07° \leq \theta_d \leq 180.00$ Venus is not "visible" at 2388 Mc. This is a somewhat surprising

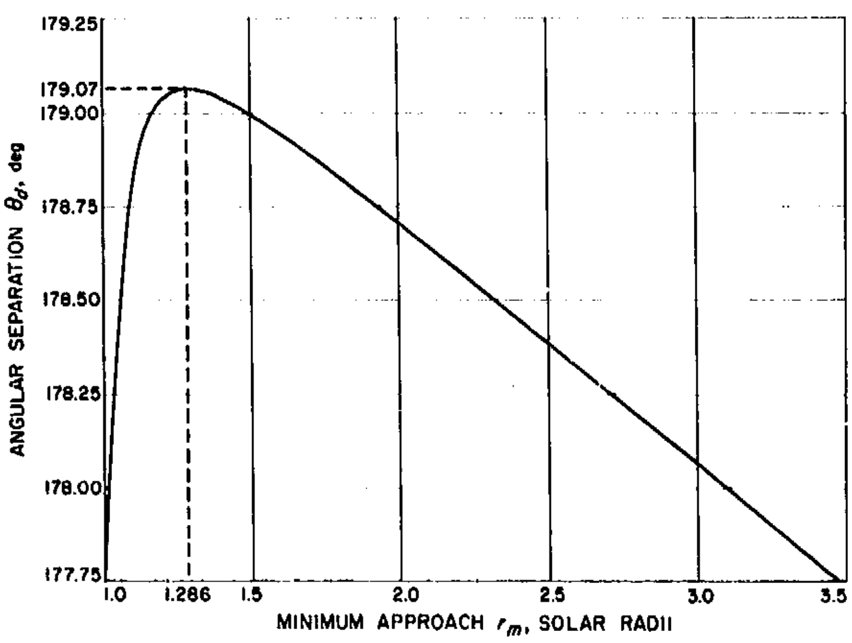


Fig. 1. Minimum approach r_m vs angular separation θ_d

result until we consider the variational principle used to find the ray equation. It can be shown that the variational principle has two local minima. This follows from the fact that we cannot imbed any single extremal in the field of the functional

$$J(\theta) = \int_{P_1}^{P_2} n(r)\left[1 + r^2\left(\frac{d\theta}{dr}\right)^2\right]^{1/2} dr$$

for $177.81° \leq \theta_d \leq 179.07°$ $[n(r)$ given by Eq. (1)$]$.

This double path will apparently cause interference patterns back at the receiver. As this problem has not been analyzed fully, we shall report on it in a future SPS article. We have, therefore, looked only at that part of the curve (Fig. 1) corresponding to $r_m \geq 1.286\, r_\odot$. The sets of ray paths corresponding to this part of the curve are minimal paths in the sense that they require less time to traverse than their counterparts.

3. Range Delay and Doppler Shift

Using Eqs. (2), (3) and (6), we programmed the problem for the digital computer. Since the integrals in Eqs. (2) and (6) have singularities at the lower limits, i.e., $r_m n(r_m) = b$, these singularities were removed for the numerical integration. This was accomplished in Eq. (2) by the change of variables.

$$\frac{[r^2 n^2(r_m) - b^2]^{1/2}}{r}\, d\rho = dr$$

i.e.,

$$r = [n^2(r_m)\rho^2 + r_m^2]^{1/2}$$

and in Eq. (6) by the change of variables

$$r(r^2 n^2 [(r_m)_k] - b_k^2)^{1/2}\, d\phi = b_k dr$$

i.e.,

$$r = \frac{(r_m)_k}{\cos\phi}$$

Both of these substitutions greatly simplified the numerical integration as the integrals became almost linear.

The results of the program are shown in Figs. 2 and 3. It is hard to compare these results with the results obtained by the simplified index in Ref. 1, Fig. 5, p. 236, and Fig. 6, p. 238. This is due to the fact that we have considered only the minimal set of paths. We do note that the range difference in Ref. 1, Fig. 5, p. 236, compares favorably with Fig. 2, at 179°, although the model used in this article causes the range difference to fall off much faster. The doppler shift in Fig. 3 is almost double that predicted in Ref. 1, Fig. 6, p. 238, at 179°. Again the index used in this article causes the effect to drop off much faster.

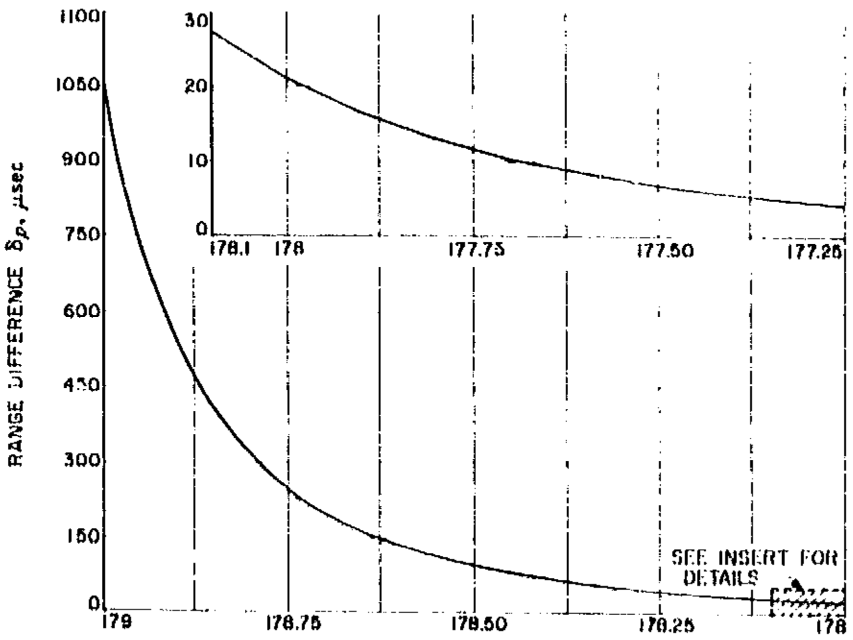


Fig. 2. Range difference $\delta\rho$ vs angular separation θ_d

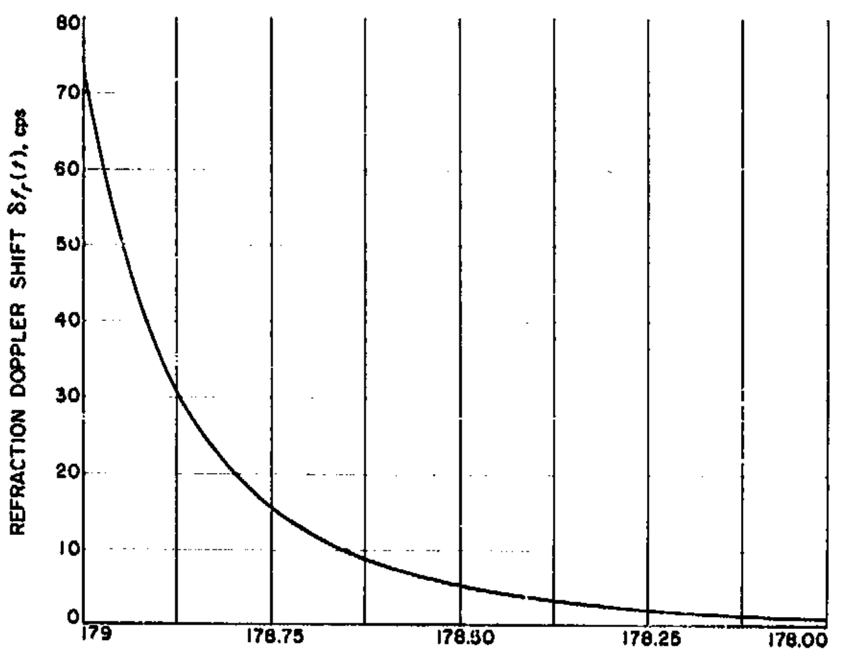


Fig. 3. Refraction doppler shift δf_r (t) vs angular separation θ_d

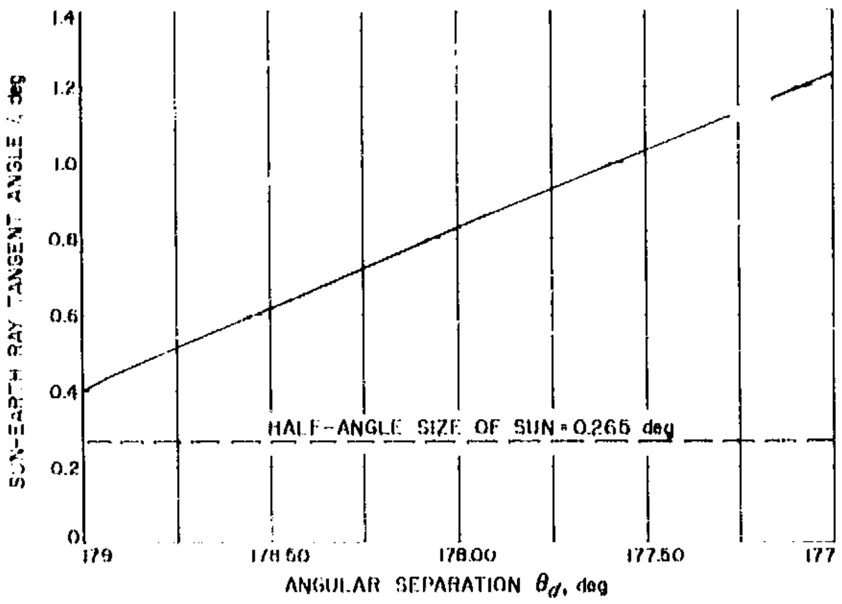


Fig. 4. Sun–Earth-ray tangent angle *i* vs angular separation θ_d

4. Conclusion

In Fig. 4 we have plotted θ_d versus the angle i formed between the radius vector to Earth and the tangent vector to the ray at the Earth. Since the half-angle size of the Sun is 0.265°, we see that using the 210-ft antenna, soon to be available, we can get close enough to the Sun for measurable effects in doppler shift and range differences.

Since we have analyzed the minimal set of ray paths only, the other set will yield larger doppler shifts and range differences, which would contribute much more to such an experiment. However, it is expected that absorption will be considerably greater for these secondary paths.

Effects still remaining to be studied in the problem include polarization variations and absorption of the radar beam due to the electron-free absorption mechanism. The study of the latter effect will require the computation of the opacity (absorption coefficient) along the ray path and the subsequent numerical integration to obtain the loss in signal strength. Investigation of the polarization effects will require a model for the solar magnetic field. Progress in these problems will be reported in subsequent issues of the SPS.

Reference

1. Reichley, P., and Muhleman, D. O., *Radar Occultations, Space Program Summaries 37-29*, Vol. IV, Jet Propulsion Laboratory, Pasadena.

XXVI. Communications Systems Research: Communication and Tracking

A. Performance of a Class of Q-Orthogonal Signals for Communication Over the Gaussian Channel

J. J. Stiffler and A. J. Viterbi

1. The Signal Set and Optimal Receiver

Consider the set of QM signals:

$$x_m^{(q)}(t)\,(m=1,2,\cdots,M;q=0,1,2,\cdots,Q-1)(0\leq t\leq T), \tag{1}$$

all of which have the same energy ST and whose normalized inner products are

$$\rho_{mn}^{qr}=\frac{1}{ST}\int_0^T x_m^{(q)}(t)\,x_n^{(r)}(t)\,dt=\begin{cases}0 \text{ for all } q,r & \text{if } m\neq n\\ \cos\dfrac{2\pi(q-r)}{Q} & \text{if } m=n\end{cases} \tag{2}$$

Examples of such signal sets are

$$x_m^{(q)}(t)=(2S)^{1/2}\sin\left(\frac{2\pi mt}{T}+\frac{2\pi q}{Q}\right)(0\leq t\leq T) \tag{3a}$$

$$x_m^{(q)}(t)=s_m(t)\sin\left(\omega_0 t+\frac{2\pi q}{Q}\right)(0\leq t\leq T) \tag{3b}$$

where $s_m(t)$ is a set of M equal energy orthogonal signals and ω_0 is a high enough frequency that the integral of the double frequency terms is negligible; or where $s_m(t)$ is a set of two-level orthogonal signals each consisting of M symbols and $\omega_0=2\pi M/T$.

Other examples are the algebraically generated polyphase code signals of Ref. 1.[1] We shall call this class of signals Q-orthogonal, since it is an obvious generalization of the class of biorthogonal signals, which corresponds to $Q=2$.

Clearly the bandwidth occupancy (Ref. 2, pp. 3–12) of any signal set in this class is:

$$W=M/T\text{ cps}, \tag{4}$$

while the transmission rate in bits per second, if all signals are equiprobable, is

$$R=\log_2(MQ)/T. \tag{5}$$

The parameter normally constrained in a digital communication system is the bandwidth-to-rate ratio

$$\frac{W}{R}=\frac{M}{\log_2 M+\log_2 Q}. \tag{6}$$

[1] Also to be presented at Symposium on Time Varying Channels, June 5 through 9, 1965, Boulder, Colorado.

The optimum receiver, when the signal set consists of MQ equal energy equiprobable signals in the presence of additive white Gaussian noise, consists of a set of correlations and a comparator and decision device which decides in favor of the signal which maximizes

$$\int_0^T y(t)\, x_m^{(q)}(t)\, dt \qquad (7)$$

where $y(t)$ is the received waveform during the given interval. It is well known that the performance of the communication system depends only on the ratio of the energy ST to the noise (one-sided) spectral density, N_0 and the $(MQ/2)$ inner products ρ_{mn}^{qr}. Thus, without loss of generality, we may assume henceforth that the signal set (3a) is used. Then

$$\int_0^T y(t)\, x_m^{(q)}(t)\, dt = (2S)^{1/2} \int_0^T y(t) \sin\left(\frac{2\pi m t}{T} + \frac{2\pi q}{Q}\right) dt$$

$$= \hat{y}_m \cos \frac{2\pi q}{Q} + \tilde{y}_m \sin \frac{2\pi q}{Q}$$

$$(m = 1, 2, \cdots, M; \quad q = 0, 1, 2, \cdots, Q - 1) \qquad (8)$$

where

$$\hat{y}_m = (2S)^{1/2} \int_0^T y(t) \sin \frac{2\pi m t}{T}\, dt \qquad (8a)$$

and

$$\tilde{y}_m = (2S)^{1/2} \int_0^T y(t) \cos \frac{2\pi m t}{T}\, t\, dt \qquad (8b)$$

or

$$\int_0^T y(t)\, x_m^{(q)}(t)\, dt = \left(\hat{y}_m^2 + \tilde{y}_m^2\right)^{1/2} \cos\left(\frac{2\pi q}{Q} - \tan^{-1} \frac{\tilde{y}_m}{\hat{y}_m}\right)$$

$$(m = 1, 2, \cdots, M; \quad q = 0, 1, 2, \cdots, Q - 1). \qquad (9)$$

Thus, it is clear that only $2M$ analog correlators, which generate the quantities of Eqs. (8a) and (8b), are necessary and that the remainder of the computation which leads to Eq. (9) can be performed by a special purpose digital computer.

2. A Sub-Optimal Receiver and Its Performance

To obtain a closed-form expression for the performance of the optimal receiver is a formidable and seemingly insurmountable task (Ref. 1). However, the following sub-optimal (but possibly nearly optimal) receiver is suggested by Eq. (8). The decision consists of two steps:

(1) Choose m so as to maximize $(\hat{y}_m^2 + \tilde{y}_m^2)^{1/2}$.

(2) With this value of m, choose q so as to maximize

$$\cos\left(\frac{2\pi q}{Q} - \tan^{-1} \frac{\tilde{y}_m}{\hat{y}_m}\right),$$

or equivalently to minimize

$$\left(\frac{2\pi q}{Q} - \tan^{-1} \frac{\tilde{y}_m}{\hat{y}_m}\right).$$

Such a receiver is easily mechanized by the parallel combination of M devices, one of which is shown in Fig. 1.

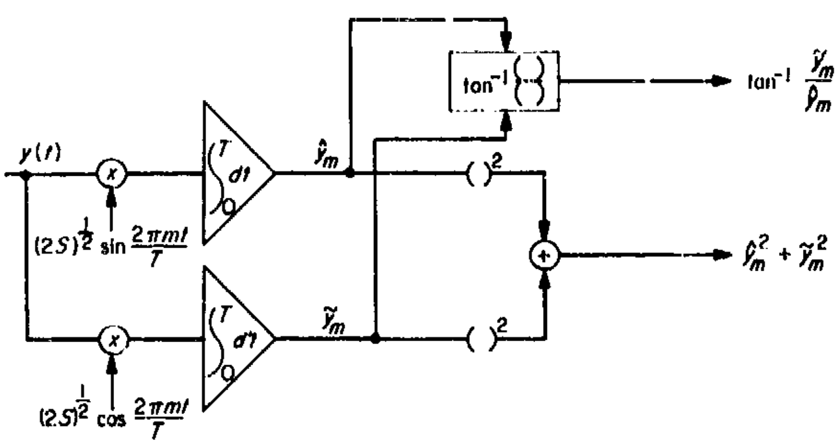


Fig. 1. An element of the sub-optimal receiver

The receiver bank resembles the optimal system for noncoherent reception. In fact, the first step in the procedure is exactly the same. The second step consists essentially of a phase estimation of the received phase angle corresponding to the largest m, comparison with each possible transmitted phase, and selection of the nearest one as the most likely. It is clear that if the signal set (3b) were used instead of (3a), the local signal inputs to the multipliers would be replaced by $s_m(t) \sin \omega_0 t$ and $s_m(t) \cos \omega_0 t$.

The error probability for this sub-optimal receiver (which, therefore, must be an upper bound on the error probability for the optimal receiver) is readily obtained. Without loss of generality, we may assume that signal

$x_1^{(0)}(t)$ was transmitted, since the signal set exhibits total symmetry. Thus, the probability of correct detection is

$$P_c = \text{Prob}\,(r_1 = \underset{m}{\text{Max}}\, r_m, \quad |\theta_1| < \pi/Q) \qquad (10)$$

where

$$r_m = (\hat{y}_m^2 + \tilde{y}_m^2)^{1/2}$$

and

$$\theta_m = \tan^{-1}\left(\frac{\tilde{y}_m}{\hat{y}_m}\right).$$

But since $x_1^{(0)}(t)$ is assumed sent and the additive noise, $n(t)$ is a zero-mean white Gaussian process of one-sided spectral density, N_0,

$$E(\hat{y}_m) = (2S)^{1/2} E \int_0^T y(t) \sin \frac{2\pi m}{T} t\, dt$$

$$= (2S)^{1/2} E \int_0^T \left[(2S)^{1/2} \sin \frac{2\pi t}{T} + n(t) \right]$$

$$\times \sin \frac{2\pi m}{T} t\, dt = \begin{cases} ST & (m = 1) \\ 0 & (m \neq 1) \end{cases}$$

$$E(\tilde{y}_m) = (2S)^{1/2} E \int_0^T y(t) \cos \frac{2\pi m}{T} t\, dt$$

$$= (2S)^{1/2} E \int_0^T \left[(2S)^{1/2} \sin \frac{2\pi t}{T} + n(t) \right]$$

$$\times \cos \frac{2\pi m}{T} t\, dt = 0$$

$$\text{Var}(\hat{y}_m) = 2S\, E \int_0^T \int_0^T n(t) \sin \frac{2\pi m t}{T} n(u) \sin \frac{2\pi m u}{T}\, du\, dt$$

$$= \int_0^T \int_0^T \frac{N_0}{2} \delta(t - u) \sin \frac{2\pi m t}{T} \sin \frac{2\pi m u}{T}\, du\, dt$$

$$= \frac{N_0 ST}{2}.$$

Similarly,

$$\text{Var}(\tilde{y}_m) = \frac{N_0 ST}{2}$$

and

$$\text{Cov}(\hat{y}_m, \tilde{y}_n) = 0 \quad (m = 1, 2, \cdots, M;\ n = 1, 2, \cdots, M).$$

Thus, the variables r_m are mutually independent, and all except r_1 have identical Rayleigh probability density functions. From Eq. (7) we have, after appropriate transformations (Ref. 3)

$$P_c = \int_{-\pi/Q}^{\pi/Q} d\theta_1 \int_0^\infty p(\theta_1, r_1)\, dr_1 \prod_{m=2}^{M} \int_0^{r_1} p(r_m)\, dr_m \qquad (11)$$

where

$$p(r_m) = \begin{cases} \dfrac{2r_m}{N_0 ST} \exp\left(-\dfrac{r_m^2}{N_0 ST}\right) & \text{for } r_m > 0 \\ 0 & \text{for } r_m < 0 \end{cases}$$

$$p(\theta_1, r_1) = \begin{cases} \dfrac{r_1}{\pi N_0 ST} \exp\left[-\dfrac{r_1^2 + (ST)^2 - 2STr_1 \cos\theta_1}{N_0 ST}\right] \\ (\text{for } r_1 \geqq 0 \text{ and } -\pi \leqq \theta_1 < \pi). \\ 0 \quad \text{otherwise} \end{cases}$$

Thus,

$$P_c = \int_{-\pi/Q}^{\pi/Q} d\theta_1 \int_0^\infty \frac{r_1}{\pi N_0 ST} \exp\left[-\frac{r_1^2 + (ST)^2 - 2STr_1\cos\theta_1}{N_0 ST}\right]\left[1 - \exp\frac{-r_1^2}{N_0 ST}\right]^{M-1} dr_1$$

$$= \frac{ST}{\pi N_0} \int_{-\pi/Q}^{\pi/Q} d\theta \int_0^\infty r \exp\left[-\left(\frac{ST}{N_0}\right)(r^2 + 1 - 2r\cos\theta)\right]\left[1 - \exp\left(-\frac{r^2 ST}{N_0}\right)\right]^{M-1} dr$$

$$= \frac{ST}{\pi N_0} \int_{-\pi/Q}^{\pi/Q} d\theta \int_0^\infty r \exp\left[\left(\frac{ST}{N_0}\right)(r^2 + 1 - 2\cos\theta)\right] \sum_{k=0}^{M-1} (-1)^k \binom{M-1}{k} \exp\left(\frac{-kr^2 ST}{N_0}\right) dr$$

$$= \sum_{k=0}^{M-1} \frac{(-1)^k \binom{M-1}{k} \exp\left(\frac{-\alpha k}{k+1}\right)}{k+1}$$

$$\times \int_{-\pi/Q}^{\pi/Q} \frac{\exp\left(\frac{-\alpha}{k+1}\right)}{2\pi} \left[1 + \left(\frac{\pi\alpha}{k+1}\right)^{1/2} \cos\theta \left\{1 + \text{erf}\left(\left[\frac{\alpha}{k+1}\right]^{1/2} \cos\theta\right)\right\} \exp\left(\frac{\alpha\cos^2\theta}{k+1}\right)\right] d\theta \qquad (12)$$

where $\alpha = ST/N_0$. The integral in the kth term of the sum is the probability distribution $P(|\theta| < \pi/Q)$ of the phase angle of a sinusoid in additive Gaussian noise for signal-to-noise ratio $\alpha/(k+1)$ (Ref. 3).

A lower bound on P_c (and consequently on upper bound on the error probability P_E) is obtained by neglecting all but the first two terms. The first term by itself is an upper bound on P_c (lower bound on P_E) for this sub-optimal receiver. In the next part of this article, we show that this is a lower bound on P_E also for the optimal receiver.

3. *Bounds on Error Probability for the Optimal Receiver*

In Part *1*, we showed that the optimum receiver selects q and m so as to maximize Eq. (6). Without loss of generality, let us assume once again that $x_1^{(0)}(t)$ was transmitted. Now using the notation of Eq. (7), we see from (6) that if

$$r_1 > \operatorname*{Max}_{m \neq 1} r_m$$

but $|\theta_1| > \pi/Q$, we shall have the correct m but the wrong q. On the other hand, if

$$r_1 < r_\beta = \operatorname*{Max}_{m \neq 1} r_m,$$

we may still be correct if $r_1 \cos\theta_1 > r_m$, and $|\theta_1| < \pi/Q$. However, if $|\theta_1| > \pi/Q$, we shall always have an error. Thus, a lower bound on the error probability for the optimum receiver is

$$P_E > \mathrm{Prob}\left(|\theta_1| > \frac{\pi}{Q}\right). \tag{13}$$

The error probability for the sub-optimal receiver is clearly an upper bound to that for the optimal receiver. Thus,

$$\int_{|\theta| > \pi/Q} p(\theta, \alpha)\, d\theta < P_E$$

$$< 1 - \sum_{k=0}^{M-1} \frac{(-1)^k}{k+1} \binom{M-1}{k} \exp\left(-\frac{\alpha k}{k+1}\right) \int_{|\theta| < \pi/Q} p\left(\theta, \frac{\alpha}{k+1}\right) d\theta$$

$$< \int_{|\theta| > \pi/Q} p(\theta, \alpha)\, d\theta + \frac{(M-1)}{2} \exp\left(-\frac{\alpha}{2}\right) \int_{|\theta| < \pi/Q} p\left(\theta, \frac{\alpha}{2}\right) d\theta \tag{14}$$

where

$$p(\theta, \gamma) =$$

$$\frac{\exp(-\gamma)}{2\pi} \{1 + (\pi\gamma)^{1/2} \cos\theta\, [1 + \mathrm{erf}(\gamma^{1/2} \cos\phi)] \exp(\gamma \cos^2\theta)\}$$

is the probability density function of the phase of the sum of a fixed phase sinusoid and Gaussian noise with $\mathrm{SNR} = \gamma$.

Finally, using a result of Arthurs and Dym (Ref. 4), we may bound the integrals in Eq. (11) by

$$\frac{1}{2}\,\mathrm{erfc}\left(\gamma^{1/2} \sin\frac{\pi}{Q}\right) < \int_{|\theta| > \pi/Q} p(\theta, \gamma)\, d\theta < \mathrm{erfc}\left(\gamma^{1/2} \sin\frac{\pi}{Q}\right)$$

so that

$$\frac{1}{2}\,\mathrm{erfc}\left(\alpha^{1/2} \sin\frac{\pi}{Q}\right) < P_E < \mathrm{erfc}\left(\alpha^{1/2} \sin\frac{\pi}{Q}\right) + \frac{(M-1)}{2}$$

$$\times \exp\left(-\frac{\alpha}{2}\right) \mathrm{erf}\left[\left(\frac{\alpha}{2}\right)^{1/2} \sin\frac{\pi}{Q}\right] \tag{15}$$

where

$$\alpha = \frac{ST}{N_0} = \frac{S}{N_0 R} \log_2(MQ)$$

and where $R = 1/T_b$ is the rate in bits per second. Letting

$$\beta = \frac{S}{N_0 R} \log_2 e,$$

$$M = \frac{1}{Q} \exp\left(\frac{k}{\beta \sin^2 \pi/Q}\right)$$

and bounding the complementary error function by

$$\left(\frac{2}{\pi}\right)^{1/2} \frac{e^{-x^2/2}}{x} \left(1 - \frac{1}{x^2}\right) < \mathrm{erfc}\left(\frac{x}{2^{1/2}}\right) < \left(\frac{2}{\pi}\right)^{1/2} \frac{e^{-x^2/2}}{x} \tag{16}$$

yields the result that

$$\frac{e^{-k}}{2(\pi k)^{1/2}} \left(1 - \frac{1}{2k}\right) < P_E < \left(\frac{e^{-k}}{(\pi k)^{1/2}}\right) + \frac{e^{k(\beta - 2)/2\beta}}{2Q}. \tag{17}$$

Consequently, in order to achieve small error probabilities, it is imperative that k be reasonably large and, since

$$M = \frac{1}{Q} \exp\left(\frac{k}{\beta \sin^2 \pi/Q}\right),$$

Q must be small compared to M. Note that if

$$\beta = \frac{S}{N_0 R} \log_e 2 > 2,$$

the upper and lower bounds on the error probability can be made arbitrarily small as k and Q are increased.

B. Another Look at the Optimum Design of Tracking Loops

R. C. Tausworthe

In 1955, Jaffe and Rechtin (Ref. 5) published the first sophisticated attempt at characterizing the optimum design of phase-locked loops. In the course of their work, they used an example which specified the transfer function of a loop best able to follow a frequency-step input insofar as minimizing transient error and phase noise are concerned. For simplicity, they assumed that the initial phase error was zero; the resulting filter function was one with one real zero and two complex poles, at a damping factor $\zeta = 0.707$, regardless of the initial frequency offset. The example was meant only to illustrate the optimization method, but since that time most systems have been designed using the parameters set by the example.

By using the same technique developed in the Jaffe–Rechtin paper, but assuming that the initial phase angle is random, a different result appears. Damping in the loop is always greater than $\zeta = 0.707$, and in all cases of practical interest, the system is *overdamped*. (Both poles lie on the negative real axis.)

Because the initial phase error is not generally known *a priori* (thus random), this latter design is one which seems to be of more practical use in most tracking applications.

1. Optimum Loops for Random Doppler Tracking

There are two sources of error during the initial acquisition of phase lock in a tracking receiver. First, there is a transient error as the system passes from its initial state to the steady-state tracking state. Second, there is phase jitter due to the presence of noise at the loop input. The technique developed by Jaffe and Rechtin was a Wiener optimization of the linearized loop, but with a constraint on the total mean-square transient error. Following this technique, the optimum loop transfer function was found to be specified by the formula

$$H_{opt}(s) = \frac{1}{[\Psi(s)]^{+}} \left[\frac{\lambda^2 \mathcal{E}\, D(s)\, D(-s)}{[\Psi(s)]^{-}} \right]_{pr}, \tag{1}$$

where

$\Psi(s) = \lambda^2 \mathcal{E}\, D(s)\, D(-s) + N_0/A^2$

$D(s) =$ doppler-phase Laplace transform

$\lambda^2 =$ Lagrange multiplier (to be evaluated)

$A^2 =$ loop input carrier power

$N_0 =$ double-sided noise spectral density

$\mathcal{E} =$ expectation operator

$[\ ]^{+} =$ left half-plane "square-root" factorization operator

$[\ ]^{-} =$ right half-plane "square-root" factorization operator

$[\ ]_{pr} = \mathcal{L}\mathcal{S}\mathcal{L}^{-1}$, the physical-realizability operator

The reader is referred to Ref. 3 or 5 for further explanation of the operators above and for the development of Eq. (1).

The optimization of interest is concerned with finding $H_{opt}(s)$ when the input doppler $d(t)$ has the form

$$d(t) = \theta_0 + \omega_0 t, \tag{2}$$

where θ_0 is a uniformly distributed phase angle, and where ω_0 is a random variable whose mean-square value is Ω_0^2. The Laplace transform of $d(t)$ is

$$D(s) = \frac{\theta_0}{s} + \frac{\omega_0}{s^2}, \tag{3}$$

and hence the expected value of $D(s)\,D(-s)$ is

$$\mathcal{E}[D(s)\,D(-s)] = -\left(\frac{\pi^2}{3}\right)\frac{1}{s^2} + (\Omega_0^2)\left(\frac{1}{s^4}\right) \tag{4}$$

The first order of business is the factorization of $\Psi(s)$:

$$[\Psi(s)]^{+} = \frac{N_0^{1/2}}{As^2}\left[s^4 + \left(\frac{\pi^2\lambda^2A^2}{3N_0}\right)s^2 + \frac{\lambda^2A^2\Omega_0^2}{N_0} \right]^{+}$$

$$= \frac{N_0^{1/2}}{As^2}\left[s^2 + \left(\frac{A^2\lambda^2\pi^2}{3N_0} + \frac{2\lambda A\Omega_0}{N_0^{1/2}}\right)^{1/2} s + \frac{\lambda A\Omega_0}{N_0^{1/2}} \right]. \tag{5}$$

It is convenient to define a quantity β^2

$$\beta^2 = \frac{\lambda A \Omega_0}{N_0^{1/2}}. \tag{6}$$

The expression for $[\Psi(s)]^-$ is obtained by substituting $-s$ for s in $[\Psi(s)]^+$. Finally, evaluation of the $[\ \ ]_{pr}$ term yields the optimum transfer function

$$H_{opt}(s) = \frac{s\left[2\beta^2 + \dfrac{\pi^2\beta^4}{3\Omega_0^2}\right]^{1/2} + \beta^2}{s^2 + \left[2\beta^2 + \dfrac{\pi^2\beta^4}{3\Omega_0}\right]^{1/2} s + \beta^2} \tag{7}$$

The corresponding loop bandwidth can be computed by integration

$$2B_L = \frac{1}{2\pi j}\int_{-j\infty}^{+j\infty} H_{opt}(s)\, H_{opt}(-s)\, ds$$

$$= \frac{\beta\left(3 + \dfrac{\beta^2\pi^2}{3\Omega_0^2}\right)}{2\left(2 + \dfrac{\beta^2\pi^2}{3\Omega_0^2}\right)^{1/2}} \tag{8}$$

from which the Lagrange multiplier λ^2 can be evaluated in terms of B_L, Ω_0, N_0, and A^2.

As we have indicated, these results are different than those contained in the Jaffe–Rechtin example. The filter they obtained, call it $H_{JR}(s)$, can be derived from the above by omitting the $\pi^2/3$ terms (i.e., by omitting those terms due the non-zero variance of θ_0):

$$H_{JR}(s) = \frac{2^{1/2}\beta s + \beta^2}{s^2 + 2^{1/2}\beta s + \beta^2} \tag{9}$$

$$2B_L = \frac{3\beta}{2(2)^{1/2}}.$$

Note that $H_{JR}(s)$ always has a damping factor of $\zeta = 0.707$, whereas the damping factor of $H_{opt}(s)$ depends on several parameters. The optimum loop can thus be specified only when Ω_0, A^2/N_0, and B_L are given.

As a further consideration, one cannot expect a very good lock-in behavior when Ω_0 is so large that the incoming carrier frequency falls outside the loop pass-band. It seems very reasonable to optimize the loop when Ω_0 lies at the edge of the pass-band[2]

$$\Omega_0 = B_L. \tag{10}$$

[2]This also corresponds to the case when a slow frequency sweep is being used, for then as Ω_0 comes within the pass-band of the loop, an optimum lock-up begins.

Under this assumption, Eq. (8) can be solved for β^2/Ω_0^2 numerically; the result is

$$\beta^2 = 1.67\Omega_0^2 = 1.67B_L^2. \tag{11}$$

By setting $\beta = 1.292B_L$ in Eq. (7) the Lagrange multiplier is removed, and the resulting optimum loop transfer function is

$$H_{opt}(s) = \frac{3.54B_L s + 1.67B_L^2}{s^2 + 3.54B_L s + 1.67B_L^2} \tag{12}$$

Specifically, $H_{opt}(s)$ has a zero at $-0.472B_L$, and poles at $-2.98B_L$ and $-0.56B_L$. Both of these poles lie on the negative real axis, indicating that the optimum system is *overdamped.*

The optimum loop filter is related to the over-all transfer function by

$$F_{opt}(s) = \frac{s\left[2\beta^2 + \dfrac{\pi^2\beta^4}{3\Omega_0^2}\right]^{1/2} + \beta^2}{AKs}$$

$$= \frac{3.54B_L s + 1.67B_L^2}{AKs} \tag{13}$$

in which K is the equivalent open-loop gain. It is usual to replace $F_{opt}(s)$ by a passive filter which degrades transient response only slightly if $AK >> B_L^2$:

$$F_{opt}(s) \approx \frac{1 + (2.12/B_L)\, s}{1 + (0.611AK/B_L^2)\, s} \tag{14}$$

2. Evaluation of Transient Error

The total transient error is given by the expression

$$\varepsilon^2 = \frac{1}{2\pi j}\int_{-j\infty}^{+j\infty}\left(-\frac{\pi^2}{3s^2} + \frac{\Omega_0^2}{s^4}\right)[1 - H_{opt}(s)]\,[1 - H_{opt}(-s)]\, ds. \tag{15}$$

With the form of the optimum filter given in Eq. (7), the transient error is

$$\varepsilon_T^2 = \frac{\Omega_0^2\left[1 + \dfrac{\pi^2\beta^2}{3\Omega_0^2}\right]}{2\beta^3\left[2 + \dfrac{\pi^2\beta^2}{3\Omega_0^2}\right]^{1/2}}, \tag{16}$$

Under the constraint $B_L = \Omega_0$, the value of β^2 is $1.67 B_L^2$, and the transient error is

$$\epsilon^2 = \frac{0.55}{B_L}. \tag{17}$$

If $H_{JR}(s)$ had been used in Eq. (15) rather than $H_{opt}(s)$, the resulting transient error would have been

$$\epsilon_{JR}^2 = \frac{\Omega_0^2 \left[1 + \frac{\pi^2 \beta_{JR}^2}{3\Omega_0^2} \right]}{2(2)^{1/2} \beta_{JR}^3}. \tag{18}$$

Given that $H_{JR}(s)$ and H_{opt} have equal bandwidths, from Eq. (9) the value of β_{JR} is $[4(2)^{1/2}/3] B_L$; the transient error thus arising by using the nonoptimal filter is

$$\epsilon_{JR}^2 = \frac{0.67}{B_L}. \tag{19}$$

This figure indicates approximately a 1-db difference in transient performance.

3. Conclusions

In many spacecraft applications, it is necessary to design a phase-tracking system whose loop bandwidth is smaller than the initial frequency uncertainty interval, and to sweep the VCO slowly to acquire lock. This indicates that the phase-locked loop should be designed optimally to acquire and track once the carrier comes within the loop pass-band. The initial phase offset of the VCO is an unknown, completely random quantity, and the loop should be designed taking this factor into account. We have specified by Eqs. (12) and (13) what the optimum loop parameters are under these circumstances. The result is an *overdamped* system, as opposed to an underdamped ($\zeta = 0.707$) system predicted by the oversimplified Jaffe–Rechtin example.

The resulting decrease in transient error using the optimum loop as compared to *misusing* the Jaffe–Rechtin example, is only about 1 db, which may nevertheless be significant in the acquisition of threshold signals.

Whenever $AK >> B_L^2$, the approximate optimum filter, given in Eq. (14) introduces a small amount of steady-state phase error, but this is generally small enough that the resulting performance of the loop is usually not degraded.

C. Some Moments Associated With Second-Order Phase-Locked Loops

H. Rumsey, Jr.

The differential equation governing the operation of a second-order phase-locked loop has the form

$$\ddot{\phi}(t) + [a + b\cos\phi(t)]\,\dot{\phi}(t) + c\sin\phi(t) = D(t), \tag{1}$$

where a, b, c are non-negative constants, ϕ is the "phase," and $D(t)$ is a certain Markovian noise process. (See *SPS 37-30* Vol. IV, pp. 262–268, for a general discussion of this problem.) We shall derive, from the Fokker–Planck equation associated with this equation, some equations which are satisfied by the expectations $E(\dot{\phi}^n e^{im\phi})$ (n, m integers). In particular, we shall show that the variances of $\dot{\phi}$ and $\sin\phi$ are connected by the simple relation

$$a\,\mathrm{Var}(\dot{\phi}) + bc\,\mathrm{Var}(\sin\phi) = \frac{D}{2}, \tag{2}$$

where D is the "mean second derivative" of the process $D(t)$, and depends only on $D(t)$.

1. Recursions for the Moments

The steady-state Fokker–Planck equation associated with Eq. (1) has the form

$$-y\frac{\partial p}{\partial \phi} + \frac{\partial}{\partial y}[(ay + by\cos\phi + c\sin\phi)\,p] + \frac{D}{2}\frac{\partial^2 p}{\partial y^2} = 0, \tag{3}$$

where we have written y in place of $\dot{\phi}$; $p = p(\phi, y)$ is the joint density function of the Markoff process (ϕ, y); D = a positive constant determined only by $D(t)$;

$$D = \lim_{\Delta t \to 0} \frac{E[\Delta D(t)]^2}{\Delta t}.$$

It is evident from the nature of the problem that $p(\phi, y)$ is a periodic function in ϕ; hence the moments $E(e^{im\phi})$ exist for all integers m. In the discussion that follows, we shall assume that the moments $E(|y|^n)$ exist and are finite for all positive integers n. If we multiply Eq. (3) by $y^n e^{im\phi}$, integrate by parts, and use this assumption and the periodicity to simplify the resulting equations, we obtain the result

$$\int_{-\infty}^{\infty}\int_{-\pi}^{\pi} e^{im\phi}\left[imy^{n+1} - ny^{n-1}(ay + by\cos\phi + c\sin\phi) + \frac{n(n-1)D}{2}y^{n-2}\right] p(\phi, y)\, d\phi\, dy = 0.$$

This can be written in the form

$$imE(y^{n+1}e^{im\phi}) = nE\left[(ay^n + by^n\cos\phi + cy^{n-1}\sin\phi)e^{im\phi}\right] - \frac{n(n-1)}{2}DE(y^{n-2}e^{im\phi}),$$

$$n = 0, 1, \cdots; \; m = 0, \pm 1, \cdots. \quad (4)$$

These equations in themselves are not sufficient to determine the moments $E(y^n e^{im\phi})$, but they do determine these moments for $n > 0$ if those for $n = 0$ are known. There is apparently no simple way of finding the moments $E(e^{im\phi})$. But, Eqs. (4) are interesting in themselves, and they can be used to derive the useful Eq. (2). This derivation follows.

2. *The Main Result*

If we let $n = 0$ and $m \neq 0$ in Eq. (4), we obtain

$$E(ye^{im\phi}) = 0. \quad (5)$$

It is evident from Eq. (1) that the behavior of the phase-locked loop is unaffected by the transformation

$$(\phi, y) \rightarrow (-\phi, -y).$$

Thus, for all $m = 0, \pm 1, \cdots$

$$E(e^{im\phi}) = E(e^{-im\phi}). \quad (6)$$

Hence, letting $n = 1$, $m = \pm 1$ in Eq. (4) and subtracting, we obtain

$$E(y^2\cos\phi) = cE(\sin^2\phi). \quad (7)$$

Now let $n = 2$, $m = 0$; this yields

$$2E(ay^2 + by^2\cos\phi) = DE(1) = D. \quad (8)$$

Combining Eqs. (7) and (8) produces

$$aE(y^2) + bcE(\sin^2\phi) = \frac{D}{2}, \quad (9)$$

which is equivalent to Eq. (2) since $E(y) = E(\sin\phi) = 0$.

Eq. (2) is important from the design standpoint, since it determines the variance of either $\sin\phi$ or ϕ when the other is known. In addition, Eq. (2) implies the upper bounds

$$\mathrm{Var}(\phi) \leq \frac{D}{2a},$$

$$\mathrm{Var}(\sin\phi) \leq \frac{D}{2abc}.$$

D. The Effects of Radio Frequency (RF) Timing Noise in Two-Way Communication Systems

W. C. Lindsey

The subject of coherent two-way communication systems is an area which is least understood both by the communication theorist and in the laboratory. Although the basic form of a two-way coherent communication system has been established and operated, emphasis must now be placed on specifying system performance, specifying optimum design trends, and seeking means of improving performance by extending present-day techniques.

In the past, the relative success with which a digital communication system performs (one-way or two-way) in the presence of noise has been prescribed on a theoretical basis in terms of a probability of error versus signal-to-noise ratio (SNR) characteristic. Associated with any communication system, which has been designed on the basis of this theoretical characteristic, is an experimentally observable operating characteristic which relates the error probability, P_E, in the transmission and reception of one binary digit to some appropriate signal-to-noise ratio (SNR) existing in the receiver. In practice, this observable operating characteristic is inferior to the theoretical characteristic and consistently takes on a shape which is offset from the theoretical (ideal) curve derived by the theoretician. This offset usually reflects the "goodness" of the particular design. Thus, the design engineer seeks means and ways of explaining this offset and ways of adjusting his design so as to narrow the offset distance.

Quite frequently, as is the case in space communications or situations where multipath is a problem, a large part of this offset is directly traceable to the failure of the system to maintain sufficiently accurate reference models of the transmitted waveforms. In particular, the inaccuracies or uncertainties in stored reference models are primarily due to a phenomenon or disturbance which may be referred to as timing noise. In practice, this means that a filter which is designed to operate as a matched filter must operate (due to timing noise) as a "randomly" mismatched filter.

This note is the first in a sequel which reports on the effects of timing noise in coherent two-way communication systems. By timing noise we mean a random disturbance which introduces uncertainties in timing, e.g., the instantaneous phase of the RF carrier or subcarrier, matched filter readout instances, etc.

In two-way coherent communication systems, there are (at least) four principal sources of timing noise which affect the number of errors present in the recovered data. These are:

(1) Phase jitter on the RF carrier due to the additive noise on the up-link.

(2) Phase jitter on the RF carrier reference due to noise on the down-link.

(3) Phase jitter on the subcarrier due to additive noise on the down-link.

(4) Readout jitter on the bit sync signal due to additive noise on the down-link.

There are probably others which produce higher-order effects.

In this note we shall consider noise sources (1) and (2), and in a subsequent article treat sources (3) and (4) and the combined effects due to all four sources.

1. Basic Model

A two-way coherent communication link (Fig. 2) is one in which an RF carrier (possibly modulated) is transmitted to the vehicle, coherently detected, filtered, and retransmitted (after appropriate modulation) to the ground receiver. The situation may be explained in the following manner: The ground transmitter (Fig. 2) emits

$$\zeta(t) = 2^{1/2} A_1 \sin[\omega_0 t + \theta_m m(t)], \tag{1}$$

and the vehicle receiver observes the doppler-shifted, noise-corrupted version

$$\Psi(t) = 2^{1/2} A_1 \sin[\omega_1 t + \theta_m m(t) + \theta_1] + n_1(t), \tag{2}$$

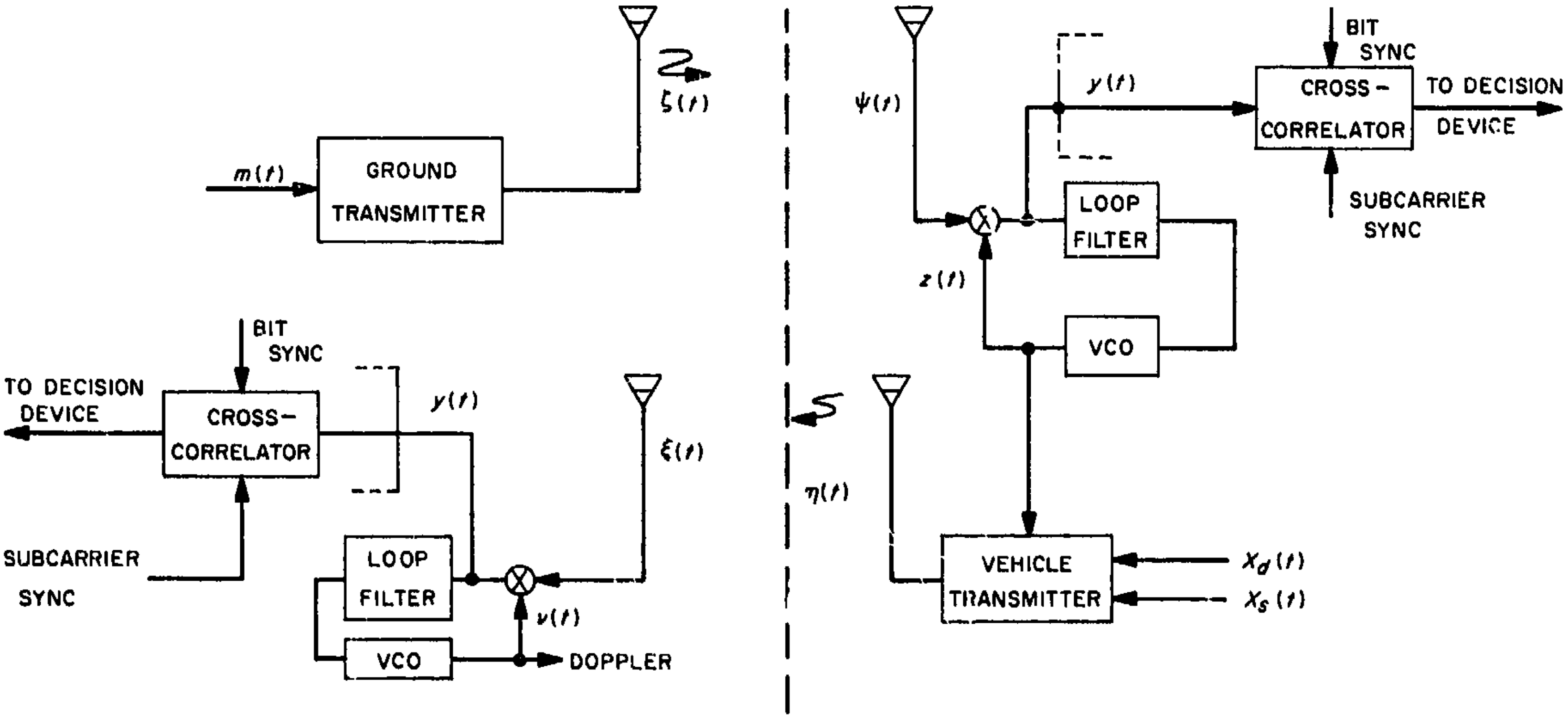


Fig. 2. Two-way communication link

where $n_1(t)$ is a narrow-band white Gaussian noise process which possesses a single-sided spectral density of N_{01} w/cps. The vehicle tracks (among other things) the carrier component in $\psi(t)$ and provides the vehicle transmitter with the carrier component. Consequently, the vehicle transmitter emits

$$\eta(t) = 2^{1/2} A_2 \sin(\omega_1 t + \theta_d X_d(t) + \theta_s X_s(t) + \hat{\theta}_1), \tag{3}$$

where $X_s(t)$ is the sync subcarrier used for deriving timing information which is needed on the ground for operating the cross-correlator and $X_d(t)$ is the data signal. In some systems, $X_s(t)$ is a PN sequence whose fundamental frequency is commensurate with that of the data bits. For phase modulation (which we are considering here), $X_d(t)$ and $X_s(t)$ are unit square waves (± 1). At the ground receiver, one observes the doppler-shifted, noise-corrupted waveform

$$\begin{aligned}\xi(t) = {} & [2^{1/2} A_2 \cos\theta_d \cos\theta_s] \sin(\omega_2 t + \hat{\theta}_1 + \theta_2) \\ & + [2^{1/2} A_2 \sin\theta_d \cos\theta_s] X_d(t) \cos(\omega_2 t + \hat{\theta}_1 + \theta_2) \\ & + [2^{1/2} A_2 \sin\theta_s \cos\theta_d] X_s(t) \cos(\omega_2 t + \hat{\theta}_1 + \theta_2) \\ & + n_2(t),\end{aligned} \tag{4}$$

where we have used a simple trigonometric expansion and neglected cross-product terms which produce third- or higher-order effects. The additive disturbance $n_2(t)$ is a narrow-band white Gaussian noise process. The power in the received carrier component, the data component, and the sync subcarrier component is, respectively,

$$\begin{aligned}P_c &= A_2^2 \cos^2\theta_d \cos^2\theta_s \\ P_d &= A_2^2 \sin^2\theta_d \cos^2\theta_s \\ P_s &= A_2^2 \sin^2\theta_s \cos^2\theta_d.\end{aligned} \tag{5}$$

Thus, the cross-modulation loss is given by

$$P_l = A_2^2 \sin^2\theta_d \sin^2\theta_s, \tag{6}$$

since the total power P must sum to A_2^2; i.e.,

$$P = A_2^2 = P_l + P_c + P_d + P_s. \tag{7}$$

The ground receiver tracks the carrier component in $\xi(t)$, which provides the receiver with the estimate $\nu(t) = 2^{1/2}\cos(\omega_2 t + \hat{\theta}_2)$. Correspondingly, this component is used to coherently demodulate $\xi(t)$. The quantity $\hat{\theta}_2$ is the phase-locked loop estimate of $\hat{\theta}_1 + \theta_2$ and, as a result of this two-way tracking and demodulation process, there will be degradation in system performance due to the RF phase-jitter, $\hat{\theta}_2$. If, in fact, $\hat{\theta}_2 = \hat{\theta}_1 + \theta_2$ this jitter is zero, and one would observe no degradation in system performance due to the RF link. (Most systems have been designed on this false assumption.)

If we multiply the observed data $\xi(t)$ with the phase-locked loop estimate $\nu(t)$ and neglect the double frequency terms, we have

$$\begin{aligned}y(t) = {} & A_2 \sin\theta_d \cos\theta_s X_d(t) \cos[\hat{\theta}_2 - (\hat{\theta}_1 + \theta_2)] \\ & + A_2 \sin\theta_s \cos\theta_d X_s(t) \cos[\hat{\theta}_2 - (\hat{\theta}_1 + \theta_2)] + n(t),\end{aligned} \tag{8}$$

where $\theta_{RF} = \hat{\theta}_2 - (\hat{\theta}_1 + \theta_2)$ is the RF phase jitter (phase error, timing noise) and $n(t)$ is a narrow-band white Gaussian noise process possessing the single-sided spectral density of N_{02} w/cps. We wish to investigate the effects which this timing-noise component and the up- and down-link additive noise components introduce into the system performance characteristic.

In order to carry out this investigation, we shall assume that the demodulator (cross-correlator) is perfectly matched; i.e., bit sync and subcarrier sync are known exactly. First, however, we point out that the statistics of the RF phase error may be shown[3] to be given by

$$p(\theta_{RF}) = \frac{I_0(|\alpha_1 + \alpha_2 \exp(j\theta_{RF})|)}{I_0(\alpha_1) I_0(\alpha_2)}; \qquad |\theta_{RF}| \leq \pi \tag{9}$$

where $I_0(x)$ is the Bessel function of zero order and imaginary argument,

$$\alpha_1 = \frac{(A_1 \cos\theta_m)^2}{N_{01} B_{L1}}$$

is the signal-to-noise ratio in the vehicle tracking loop bandwidth B_{L1}, and

$$\alpha_2 = \frac{(A_2 \cos\theta_s \cos\theta_d)^2}{N_{02} B_{L2}}$$

is the signal-to-noise ratio in the ground receiver tracking bandwidth B_{L2}.

[3] "Two-Way Doppler and Phase Measurements in Communication Networks," by W. C. Lindsey (to be published).

Without presenting the complicated and tedious details of the derivation, it is possible to show that the probability that the demodulator (cross-correlator) errs is given by

$$P_E(n) = \frac{1}{2}\left[1 - \left(\frac{8R_n}{\pi}\right)^{1/2} \sum_{k=0}^{\infty} \exp(-R_n)(-1)^k \epsilon_k b_{2k+1}(n) \times (1-4k^2)^{-1} I_k(R_n)\right], \qquad (10)$$

where

$$b_k(n) = \prod_{i=1}^{n} \frac{I_k(\alpha_i)}{I_0(\alpha_i)}; \qquad n = 1, 2.$$

$$\epsilon_k = \begin{cases} 1; & \text{if } k = 0 \\ 2; & \text{if } k > 0 \end{cases}$$

$$R_n = S_n T_n / N_{0n}$$

or

$$R_1 = (A_1 \cos\theta_m)^2 T_1/N_{01} = S_1 T_1/N_{01}$$

= data signal-to-noise ratio in vehicle

$$R_2 = (A_2 \sin\theta_d \cos\theta_s)^2 T_2/N_{02} = S_2 T_2/N_{02}$$

= data signal-to-noise ratio on ground

If $n = 1$ in Eq. (10), we have the performance of the demodulator in the vehicle, while letting $n = 2$ yields the probability that the demodulator will make an error on the ground. If we fix the signal-to-noise ratios in the tracking loops and let R_n approach infinity, it is easy to show that the system is plagued with an irreducible error probability. This irreducible error rate is given, as a function of the signal-to-noise ratios in the tracking loops, by

$$P_E(n) = \frac{1}{2}\left[1 - \frac{2}{\pi} \sum_{k=0}^{\infty} (-1)^k \frac{\epsilon_k b_{2k+1}(n)}{1-4k^2}\right] \qquad (11)$$

A convenient check on the result is to let $\alpha_1 = \alpha_2$ approach infinity in Eq. (10). This corresponds to zero RF phase error, i.e., perfect measurement, and we have (Ref. 6)

$$P_E(n) = \frac{1}{2\pi}\int_{(2R_n)^{1/2}}^{\infty} \exp(-y^2/2)\, dy$$

which checks with previous results. For various values of R_n, Eq. (2) is plotted in Fig. 3 for the case where $\alpha_1 = \alpha_2 = \alpha$ and for $n = 1$ and 2. The dashed curves ($n = 2$) represent the error probability versus the parameter R_n when both the up-link and down-link are equally

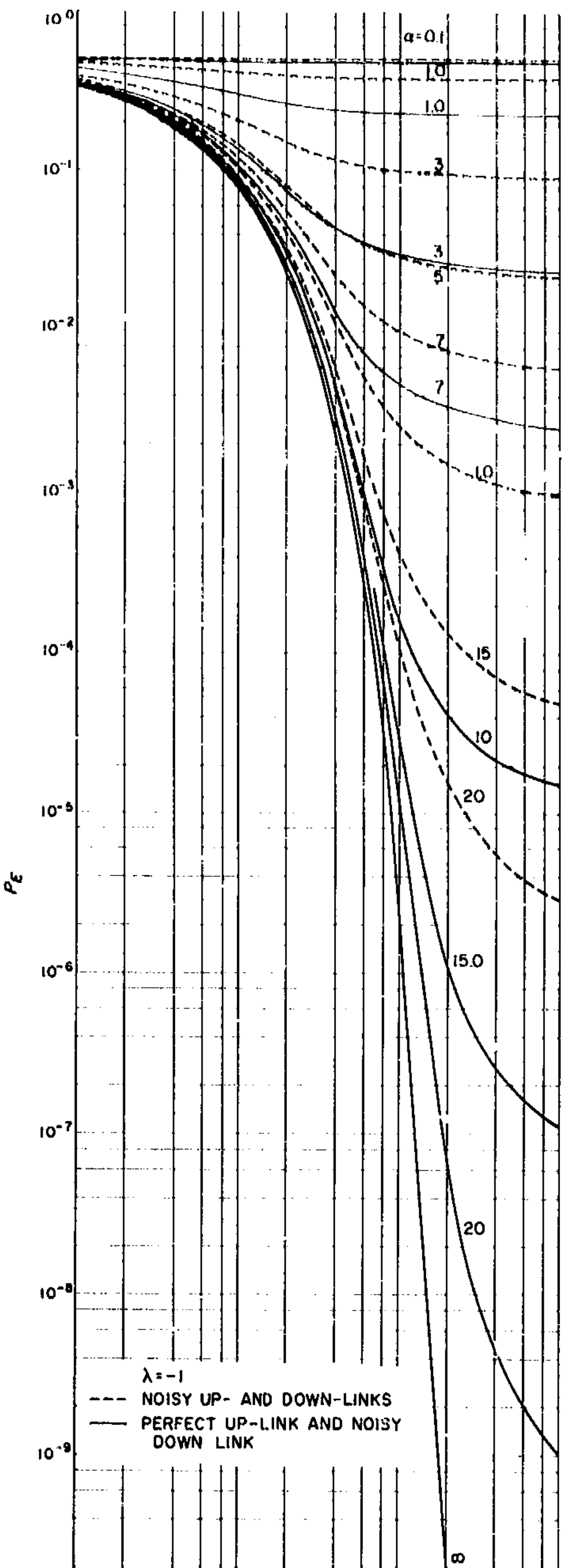


Fig. 3. Bit-error probability versus signal-to-noise ratio

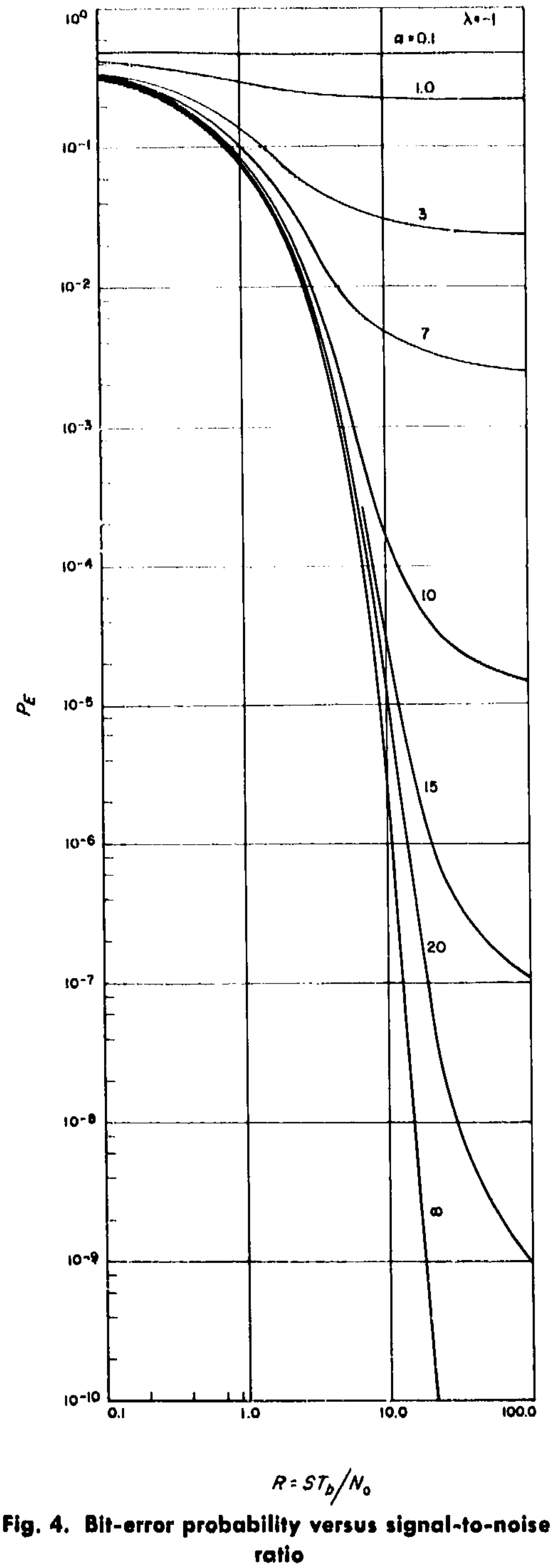


Fig. 4. Bit-error probability versus signal-to-noise ratio

noisy. The solid curves indicate the performance of the ground receiver when the up-link is non-noisy. Fig. 4 illustrates the performance of the vehicle receiver. Notice how the up-link timing noise affects the performance of the ground receiver and how the timing noise on the up- and down-links degrades over-all performance.

As an example of the use of these results, consider the situation where a two-way system has been designed to operate on the basis of zero RF-timing error and the $P_E(2) = 10^{-3}$. Suppose further that $\alpha_1 = \alpha_2 = \alpha = 15$. Entering Fig. 3 and using these values, we find that, to maintain an error rate of 10^{-3}, the signal-to-noise ratio in the data channel would have to be increased by approximately 2.4 db over what would be necessary if system timing references were perfect. This is a significant loss when one considers the cost per db of increasing the system signal-to-noise ratio. For one-way operation, the loss is approximately 0.6 db.

From a practical standpoint, it would be interesting to optimize system performance by varying the ground system or vehicle system modulation index so as to minimize $P_E(n)$ at both link ends. Further, it would be interesting to check these results with those obtained from experiment. Also of future interest is to determine optimum receiver structures, to extend the present analysis to "n-step" networks, to consider the effects of timing noise sources (3) and (4) and to consider the combined effects of all four sources.[4]

E. On the *A Priori* Information in Multi-Stage Estimation Problems

T. Nishimura

The optimal filter which has been introduced by Kalman (Refs. 7, 8) into the field of system theory yields the minimum variance estimate of states of linear systems, which are contaminated by white Gaussian noises, when a set of sequential observations is carried out. The basic feature of this filter is that the estimate of states is updated by a sequence of observations so as to minimize its

[4] Preliminary results which pertain to timing noise sources (3) and (4) are given in "The Detection of PSK Signals Using a Noisy Phase Reference," by W. C. Lindsey, *Proceedings of the National Telemetering Conference*, First Edition, pp. 50–53, April 1965.

variance, or equivalently, if noise is Gaussian, to maximize the conditional probability density of the current states after having a set of observations.

The original theory assumes an *a priori* information on the initial states and their variances. However, there are cases when this *a priori* information may contain certain errors, or some information may not be available at the beginning of the estimating process.

Soong (Ref. 9) examined the effect of errors in the *a priori* data on the *a posteriori* variance of estimates for the case of single-stage observation and derived the deviations of the calculated and actual variances from the true minimum variance.

The same problem is studied in this report for the multi-stage case, i.e., for the case when sequential observations and estimations are performed on the state. Then this optimal filter, which can be synthesized in the form of a program on the digital computer, thus assuming a form of digital filter, is connected to the existing analog filter in tandem. By means of this tandem configuration of analog filter and digital filter, shown in Fig. 5, the detection of states as well as signals can be speeded up considerably, while the fluctuation of the final observation is confined within a small allowance with constantly decreasing variance. A phase-locked loop receiver is used as an example of the analog filter. The practical importance of the extended results on the effect of errors in the *a priori* information is demonstrated in this specific example.

1. Fundamental Equations

The system treated is a discrete-time linear system with the discrete interval T; its state vector representations are described by the following difference equations:

$$x(k+1) = \Phi(k)\,x(k) + G(k)\,w(k). \tag{1}$$

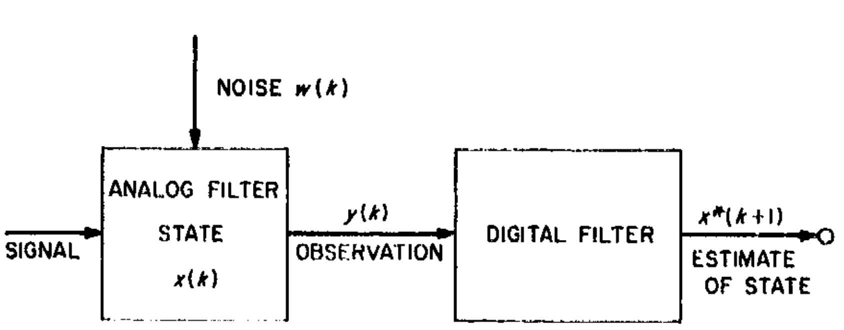


Fig. 5. Tandem configuration of analog and digital filter

The observation is

$$y(k) = M(k)\,x(k), \tag{2}$$

where

x = a column vector of dimension n_x, the state vector of the system

w = a column vector of dimension n_w, the noise input to the system

y = a column vector of dimension n_y, the observation of the system output

Φ = an $n_x \times n_x$ matrix, the state transition matrix

G = an $n_x \times n_w$ matrix, the noise coefficient matrix

M = an $n_y \times n_x$ matrix, the observation matrix

The problem is now to find an optimum estimate of $x(k+1)$ based on the observation $y(0), y(1), \cdots, y(k)$. One solution to this problem employing a Bayesian approach is given by Ho (Ref. 10), which agrees with Kalman's results when the noise is Gaussian. In this solution, the conditional density function

$$p\,[x(k+1)\,|\,y(0), y(1), \cdots, y(k)]$$

is derived by means of the Bayes rule, observing that both x and y are Gaussian when w is Gaussian. Then the conditional expectation of a quadratic loss function

$$\langle (x - x^*)^T (x - x^*) \rangle$$

is derived and minimized, thus yielding the following equations for the optimum estimator x^* of the state vector x. Here, $\langle\ \rangle$ denotes the expected-value operator. Let Q be the covariance matrix of the noise w (assuming its mean is zero),

$$\langle w \rangle = 0 \tag{3}$$

$$\langle w \cdot w^T \rangle = Q, \tag{4}$$

and let $P(k)$ be a square matrix of dimension n_x which represents a covariance function of the estimator:

$$P(k) = \langle [x(k) - x^*(k)]\,[x(k) - x^*(k)]^T \rangle. \tag{5}$$

Then, given *a priori* information on initial conditions $x^*(0)$ and $P(0)$, the optimal estimator $x^*(k)$ of the state vector $x(k)$ is recursively defined by (Refs. 7, 10)

$$x^*(k+1) = \Phi(k)\,x^*(k) + K(k)\,[y(k) - M(k)\,x^*(k)] \tag{6}$$

$$K(k) = \Phi(k)\,P(k)\,M^T(k)\,[M(k)\,P(k)\,M^T(k)]^{-1} \tag{7}$$

and

$$P(k+1) = [\Phi(k) - K(k)M(k)]P(k) \times [\Phi(k) - K(k)M(k)]^T + G(k)QG^T(k). \tag{8}$$

The digital filter of Fig. 5 is represented by Eqs. (6) through (8); its matrix block diagram is drawn in Fig. 6.

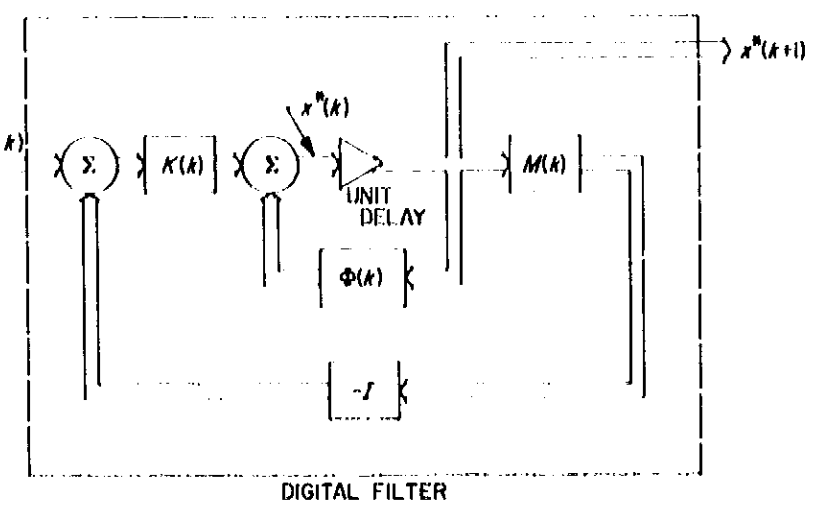


Fig. 6. Matrix block diagram of digital filter

Combining the two equations,

$$P(k+1) = \Phi(k)\{P(k) - P(k)M^T(k)[M(k)P(k)M^T(k)]^{-1} \times M(k)P(k)\}\Phi^T(k) + G(k)QG^T(k). \tag{9}$$

These recurrence relations yield the optimum estimate of the state vectors sequentially, provided the *a priori* information on the initial conditions $x^*(0)$ and $P(0)$ is available. However, if such *a priori* information is incorrect, the subsequent $x^*(k)$ and $P(k)$ will be inevitably in error. Therefore, an analysis becomes necessary on the behavior of the error caused by the incorrect *a priori* information, since it is highly probable that in practical circumstances the *a priori* information is not sufficiently correct, or that it is not available at all.

As mentioned in the introduction, Soong considered this problem for the single-stage case (that is, the case when single estimate is performed following single observation of the state).

This article extends the analysis to the multi-stage case, (that is, the case when sequential estimates are carried out following sequential observations on the states).

2. Derivation of Error Matrices

Adopting Soong's notations, three different kinds of covariance matrices are introduced, namely P_o, P_c and P_a:

P_o = optimum covariance matrix, when correct *a priori* information is employed

P_c = calculated covariance matrix derived by Eq. (9), when incorrect information is employed for the initial condition

P_a = actual covariance matrix derived by Eq. (5), when the estimate is computed by using the calculated covariance matrix with the incorrect *a priori* data

In the subsequent analysis, a subscript o is attached to the optimum P and K described by Eqs. (7) and (8) in order to discriminate them from other similar quantities such as P_c or K_c.

Picking up an arbitrary instant k, the error function at the next instant, i.e., the $(k+1)$th instant, is derived as a function of the error function at the kth instant. Then, by means of induction, a useful property of the error function shall be proved over all the instants, starting from the 0th instant.

Since there are three kinds of covariance matrices, three error functions are defined as their mutual differences. They are:

$$E_{ca}(k) = P_c(k) - P_a(k) \tag{10}$$

$$E_{co}(k) = P_c(k) - P_o(k) \tag{11}$$

$$E_{ao}(k) = P_a(k) - P_o(k). \tag{12}$$

It is known that $P_o(k)$, which is an optimum covariance matrix defined by Eq. (5), is non-negative definite for all $k \geq 0$ (Ref 7). Further, we assume that MP_oM^T and MP_cM^T are invertible for the sake of simplicity throughout this article.

a. Calculation of E_{ca}. We notice that the calculated covariance matrix P_c is derived by recurrence formulas Eqs. (7) and (8) starting from the given initial condition $P_c(0)$, and it is completely independent of x, x^*, or y. On the other hand, the actual covariance matrix is derived from Eq. (5) by definition, in which the estimate x^* is computed by Eq. (6) using the calculated K and P in it. Hence, it is dependent on x, x^*, and y as well as on P_c and K_c.

In order to simplify the description, the time (k) shall be dropped from all the quantities concerned except when it is more advantageous to indicate it explicitly in the subsequent analysis. Hence, from Eqs. (7) and (8)

$$P_c(k+1) = (\Phi - K_cM)\,P_c\,(\Phi - K_cM)^T + GQG^T, \tag{13}$$

where

$$K_c = \Phi P_cM^T\,[MP_cM^T]^{-1}. \tag{14}$$

On the other hand, $P_a(k+1)$ is described as follows: denoting the actual, nonoptimum estimate as x_a^* when it is derived using K_c and P_c in Eqs. (6) and (7), we have

$$\begin{aligned} P_a(k+1) &= \langle [x_a^*(k+1) - x(k+1)]\,[x_a^*(k+1) \\ &\quad - x(k+1)]^T \rangle \\ &= \langle \{ x_a^*(k+1) - x^*(k+1) \\ &\quad + [x^*(k+1) - x(k+1)]\}\,\{[x_a^*(k+1) \\ &\quad - x^*(k+1)] + [x^*(k+1) - x(k+1)]\}^T \rangle, \end{aligned} \tag{15}$$

where $x(k+1)$ and $x^*(k+1)$ are given by Eqs. (1) and (6), respectively, while $x_a^*(k+1)$ is described as

$$x_a^*(k+1) = \Phi x_a^* + K_c(y - Mx_a^*). \tag{16}$$

Then, from Eqs. (2), (6), and (16)

$$\begin{aligned} x_a^*(k+1) - x^*(k+1) &= (\Phi - K_cM)\,(x_a^* - x^*) \\ &\quad + (K_o - K_c)\,M\,(x^* - x)\,. \end{aligned} \tag{17}$$

Also, Eqs. (1), (2), and (6) yield

$$x^*(k+1) - x(k+1) = (\Phi - K_oM)\,(x^* - x) - Gw. \tag{18}$$

Substitute Eqs. (17) and (18) into Eq. (15). Note that $x(k)$ and $x^*(k)$ are independent from $w(k)$ because w is a white noise. Then Eq. (15) can be reduced to the following simple form after a lengthy manipulation of matrices:

$$P_a(k+1) = (\Phi - K_cM)\,P_a\,(\Phi - K_cM)^T + GQG^T. \tag{19}$$

The difference between Eqs. (13) and (19) yields the desired recurrence relationship for E_{ca}.

$$\begin{aligned} E_{ca}(k+1) &= (\Phi - K_cM)\,[P_c(k) - P_a(k)]\,(\Phi - K_cM)^T \\ &= (\Phi - K_cM)\,E_{ca}(k)\,(\Phi - K_cM)^T\,. \end{aligned} \tag{20}$$

b. Calculation of E_{co}. Since $P_c(k+1)$ and $P_o(k+1)$ are given by Eqs. (13) and (8), respectively, their difference is

$$\begin{aligned} E_{co}(k+1) &= P_c(k+1) - P_o(k+1) \\ &= (\Phi - K_cM)\,P_c\,(\Phi - K_cM)^T \\ &\quad - (\Phi - K_oM)\,P_o\,(\Phi - K_oM)^T \\ &= (\Phi - K_cM)\,E_{co}\,(\Phi - K_cM)^T \\ &\quad + (\Phi - K_oM)\,[E_{co}M^T\bar{P}_c^{-1}\bar{P}_o\bar{P}_c^{-1}ME_{co} \\ &\quad - P_cM^T\bar{P}_c^{-1}ME_{co} - E_{co}M^T\bar{P}_c^{-1}MP_c] \\ &\quad \times (\Phi - K_oM)^T \end{aligned} \tag{21}$$

where

$$K_c = K_o + (\Phi - K_oM)\,E_{co}M^T\bar{P}_c^{-1}. \tag{22}$$

The bar above indicates the multiplication of M to the left and M^T to the right of the underlying quantity, e.g.,

$$\bar{P}_c = MP_cM^T. \tag{23}$$

When Eq. (21) is expanded and arranged together, it can be finally reduced to the following concise form:

$$\begin{aligned} E_{co}(k+1) &= (\Phi - K_oM)\,[E_{co}(k) - E_{co}(k)\,M^T\bar{P}_c(k)^{-1} \\ &\quad \times ME_{co}(k)]\,(\Phi - K_oM)^T. \end{aligned} \tag{24}$$

c. Calculation of E_{ao}. $E_{ao}(k+1)$ is obtained either by taking the difference between $P_a(k+1)$ and $P_o(k+1)$ of Eqs. (19) and (8),

$$\begin{aligned} E_{ao}(k+1) &= P_a(k+1) - P_o(k+1) \\ &= (\Phi - K_cM)\,P_a\,(\Phi - K_cM)^T \\ &\quad - (\Phi - K_oM)\,P_o\,(\Phi - K_oM)^T \end{aligned} \tag{25}$$

or by subtracting E_{ca} from E_{co},

$$E_{ao}(k+1) = E_{co}(k+1) - E_{ca}(k+1), \tag{26}$$

where $E_{co}(k+1)$ and $E_{ca}(k+1)$ are given by Eqs. (24) and (20), respectively.

Either method leads to the following result, which is represented as a sum of two symmetric matrices after certain matrix operations:

$$\begin{aligned} E_{ao}(k+1) &= (\Phi - K_cM)\,E_{ao}(k)\,(\Phi - K_cM)^T \\ &\quad + (\Phi - K_oM)\,[E_{co}(k)\,M^T\bar{P}_c(k)^{-1} \\ &\quad \times \bar{P}_o(k)\,\bar{P}_c(k)^{-1}ME_{co}(k)]\,(\Phi - K_oM)^T. \end{aligned} \tag{27}$$

3. Evaluation of Results

We observe that $E_{ea}(k+1)$, $E_{eo}(k+1)$, and $E_{ao}(k+1)$ derived in the previous part of this article are all symmetric, square matrices. For the purpose of evaluating these recurrence formulas, the following three facts concerning real matrices are useful:

(1a) Given an arbitrary matrix B, $B \cdot B^T$ is non-negative definite.

(1b) Conversely, if H is a non-negative definite symmetric matrix, it can be factored into the form $H = B \cdot B^T$.

(2) If H_1 and H_2 are symmetric and non-negative definite matrices, then $H_1 + H_2$ is non-negative definite.

(3) If H is a symmetric non-negative definite matrix, then every diagonal component of H is non-negative.

Then the following three theorems are presented and proved as the principal results of this report.

Theorem 1. If $E_{ea}(k)$ is non-negative definite, then $E_{ea}(k+1)$ is non-negative definite.

Proof. If $E_{ea}(k)$ is non-negative definite, by (1b), it can be factored as

$$E_{ea}(k) = B \cdot B^T. \tag{28}$$

Then by letting

$$C = (\Phi - K_e M) B \tag{29}$$

and referring to Eq. (20), $E_{ea}(k+1)$ can be factored as

$$E_{ea}(k+1) = C \cdot C^T. \tag{30}$$

Thus, it becomes clear that $E_{ea}(k+1)$ is non-negative definite according to (1a).

Theorem 2. If $E_{eo}(k)$ is non-negative definite, then $E_{eo}(k+1)$ is non-negative definite.

Proof. When $E_{eo}(k)$ is positive definite, the content of the parentheses in the middle of Eq. (24) can be rewritten as a sum of two symmetric matrices as follows, assuming that $ME_{eo}(k)M^T$ is also positive definite, and hence invertible.

$$\begin{aligned} E_{eo}(k) - E_{eo}(k)M^T\bar{P}_e(k)^{-1}ME_{eo}(k) &= E_{eo}(k)M^T[\bar{E}_{eo}(k) \\ &\quad + \bar{E}_{eo}(k)\bar{P}_o(k)^{-1}\bar{E}_{eo}(k)]^{-1}ME_{eo}(k) \\ &\quad + [I - E_{eo}(k)M^T\bar{E}_{eo}(k)^{-1}M]E_{eo}(k) \\ &\quad \times [I - E_{eo}(k)M^T\bar{E}_{eo}(k)^{-1}M]^T. \end{aligned} \tag{31}$$

Since the inverse of a positive definite matrix is also positive definite, the positive definiteness of the quantity in the right-hand side of the previous equation can be proved with the help of (2) and using the result in the proof of Theorem 1. Therefore, the non-negative definiteness of $E_{eo}(k+1)$ in Eq. (24) is claimed, referring again to the proof in Theorem 1.

When $E_{eo}(k)$ is non-negative definite, $E_{eo}(k) + \epsilon I$ ($\epsilon > 0$) is positive definite. Substituting this sum for $E_{eo}(k)$ in Eq. (31), the non-negative definiteness of the left-hand side of Eq. (31) can be proved as a limiting case as $\epsilon \to 0$. Thus, the statement at the beginning is verified.

Theorem 3. If $E_{ao}(k)$ is non-negative definite, then $E_{ao}(k+1)$ is non-negative definite.

Proof. In Eq. (27), the second term is always non-negative definite because the covariance matrix P_o is non-negative definite, so that a proof similar to the one in Theorem 1 can be applied. Also, the first term is non-negative definite when $E_{ao}(k)$ is non-negative definite. Therefore, the above theorem is verified, referring again to (2).

Finally, an important conclusion is drawn by applying induction to these results; namely, the error matrices $E_{ea}(k)$ and $E_{eo}(k)$ are non-negative definite for every $k \geq 1$ whenever their respective initial conditions $E_{ea}(0)$ and $E_{eo}(0)$ are non-negative definite, and $E_{ao}(k)$ is always non-negative definite because $E_{ao}(0)$ is non-negative definite as deduced from the definition of $P_o(0)$.

Consequently, all the diagonal components of $E_{ea}(k)$, $E_{eo}(k)$ are non-negative (for all $k \geq 0$) by Eq. (3) when $E_{ea}(0)$ and $E_{eo}(0)$ are, respectively, non-negative definite. The diagonal components of $E_{ao}(k)$ are non-negative for all $k \geq 0$ without any condition.

This result will be very useful in practical applications.

Especially, the non-negativeness of $e_{ea,ii}(k)$, which represents a diagonal component of $E_{ea}(k)$, implies

$$p_{eii}(k) \geq p_{aii}(k) \geq 0, \tag{32}$$

where $p_{eii}(k)$ and $p_{aii}(k)$ denote diagonal components of $P_e(k)$ and $P_a(k)$, respectively. It is clear that $p_{aii}(k)$ is non-negative because the covariance matrix $P_a(k)$ is non-negative definite.

Therefore, the actual variance of estimates (which is nonaccessible) based on an incorrect *a priori* information is bounded by the calculated variance (which is accessible) if the initial error matrix $E_{ea}(0)$ is so selected as to be non-negative definite.

4. Application to the Phase-Locked Loop System

For the purpose of demonstrating an application of the theory developed so far to the tandem configuration of analog and digital filter, a phase-locked loop is taken as an example of such an analog filter.

When the VCO is represented by an integrator and the combined loop gain by a, the feedback loop configuration of the phase-locked loop becomes as shown in Fig. 7. The slope of the error function is represented by A. We assume that the incoming noise is white and Gaussian with density N_o (one-sided).

When the input to the loop is a phase step, it can be considered as a state of an expanded system constructed by adding a constant voltage to the original system at its input. This input state is described as x_2 and the output of the phase-locked loop as x_1. Then the state vector at $t = (k+1)T$ is described as a linear function of the state at $t = kT$.

$$x_1(k+1) = e^{-bT}x_1(k) + (1-e^{-bT})\,x_2(k) + \frac{(1-e^{-bT})}{A}\,w_1(k) \tag{33}$$

$$x_2(k+1) = x_2(k) \tag{34}$$

where the equivalent loop gain is given by

$$b = aA = \frac{1}{\tau} \tag{35}$$

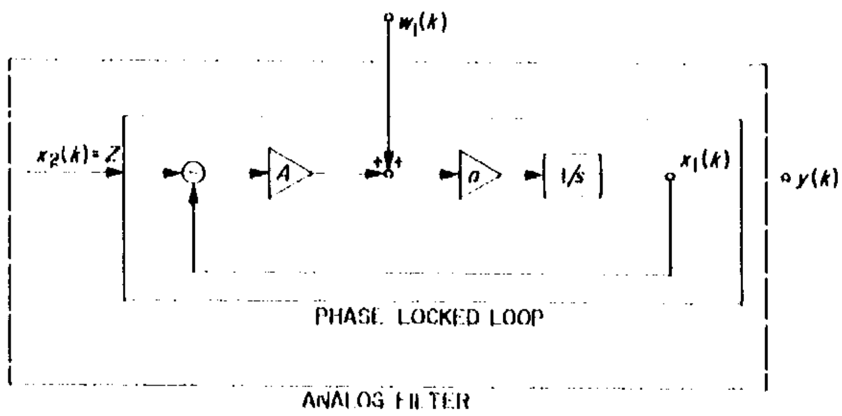


Fig. 7. Analog filter (phase-locked loop)

and the τ is the time constant of the loop. Eqs. (33) and (34) constitute the analog filter in Fig. 5. Then the corresponding vectors and matrices in Eq. (1) become

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{36}$$

$$\Phi = \begin{bmatrix} e^{-bT} & 1-e^{-bT} \\ 0 & 1 \end{bmatrix} \tag{37}$$

$$G = \begin{bmatrix} (1-e^{-bT})/A & 0 \\ 0 & 0 \end{bmatrix} \tag{38}$$

$$w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}. \tag{39}$$

The observation y is identical to the phase output x_1.

$$y(k) = x_1(k). \tag{40}$$

Hence, M of Eq. (2) becomes a row matrix

$$M = [1, 0]. \tag{41}$$

The direct substitution of these matrices into Eq. (6) yields the optimum estimator as

$$x^*(k+1) = \Phi\hat{x}^*(k), \tag{42}$$

where $\hat{x}^*(k)$ is given by

$$\hat{x}_1^*(k) = y(k) \tag{43}$$

$$\hat{x}_2^*(k) = x_2^*(k) + \frac{p_{12}(k)}{p_{11}(k)}\,[y(k) - x_1^*(k)]. \tag{44}$$

The recurrence formulas for the P-matrix are

$$p_{11}(k+1) = \frac{p_{11}(k)\,p_{22}(k) - p_{12}^2(k)}{p_{11}(k)}\,\phi_{12}^2 + \sigma_u^2 \tag{45}$$

$$p_{12}(k+1) = \frac{p_{11}(k)\,p_{22}(k) - p_{12}^2(k)}{p_{11}(k)}\,\phi_{12} = p_{21}(k+1) \tag{46}$$

$$p_{22}(k+1) = \frac{p_{11}(k)\,p_{22}(k) - p_{12}^2(k)}{p_{11}(k)} \tag{47}$$

where

$$\phi_{12} = 1 - e^{-bT} \tag{48}$$

and

$$\sigma_u^2 = \frac{(1 - e^{-bT})^2\,N_o}{2A^2T}. \tag{49}$$

Eqs. (42) through (47) specify the optimal digital filter shown in Figs. 5 and 6.

The principal objective of this digital filter is, of course, to find the optimum estimate of the state x_2 at every instant of time, by suppressing the effect of noise disturbance as much as possible.

Assuming that the system is at rest for $t < 0$, and that the *a priori* information on $x^*(0)$ and $P(0)$ is correct,

$$x_1^*(0) = 0 \tag{50}$$

$$x_2^*(0) = Z \tag{51}$$

$$p_{11}(0) = \sigma_u^2 \tag{52}$$

$$p_{12}(0) = p_{21}(0) = 0 \tag{53}$$

$$p_{22}(0) = 0, \tag{54}$$

where Z is the amplitude of the actual step input. Then the optimum estimators become

$$x_1^*(k+1) = e^{-bT}y(k) + Z(1 - e^{-bT}) \tag{55}$$

$$x_2^*(k+1) = Z \tag{56}$$

and

$$p_{11}(k+1) = \sigma_u^2 \tag{57}$$

$$p_{12}(k+1) = p_{21}(k+1) = p_{22}(k+1) = 0 \tag{58}$$

for all $k \geq 0$.

However, the *a priori* information may be incorrect, especially $x_2^*(0)$ and $p_{22}(0)$. (If they are correct, there is no need to proceed on the estimation.) Therefore, we inevitably start the estimation process of Eqs. (7) through (9) assigning a set of initial conditions which are not the optimum ones. Fortunately, the discussion in Part 3 of this article assures us that the actual variance of the estimate is bounded by the calculated variance if $E_{ca}(0)$ is selected to be non-negative definite. Hence, it becomes possible to study the real-time behavior of the digital filter using the P_c matrix as an upper bound for the actual covariance matrix.

The behavior of P_c as a function of time is obtained as a solution of three nonlinear difference equations described by Eqs. (45) through (47) starting from the assumed *a priori* information $P_c(0)$. Certain manipulations of Eqs. (45) through (47) yield a nonlinear difference equation of p_{c22} alone which plays the most important role among the components of the P_c matrix in this analysis.

$$\phi_{12}^2 p_{c22}(k+1)\,p_{c22}(k) + \sigma_u^2\,[p_{c22}(k+1) - p_{c22}(k)] = 0. \tag{59}$$

When the discrete interval T is very short, the above difference equation can be replaced by the corresponding differential equation. Further, we assume that $T\,(dp_{c22}/dt)$ is much smaller than p_{c22} when T is a sufficiently short interval. This assumption is acceptable since we are more concerned with the asymptotic behavior of p_{c22} for large t, where $p_{c22}(t)$ is also small as explained later. In this case, the above difference equation is reduced to the following Riccati type differential equation:

$$\frac{dp_{c22}}{dt} = -\frac{\alpha}{2\tau}\,p_{c22}^2, \tag{60}$$

where

$$\alpha = \frac{2\tau\phi_{12}^2}{\sigma_u^2 T} = \frac{A^2}{N_o B_L} \tag{61}$$

Thus, α is the signal-to-noise ratio in terms of the loop bandwidth B_L, where

$$B_L = \frac{b}{4} = \frac{1}{4\tau}. \tag{62}$$

The solution of this differential equation is

$$p_{e22}(t) = \frac{p_{e22}(0)}{\alpha p_{e22}(0)(t/2\tau) + 1}. \tag{63}$$

We observe that $p_{e22}(t)$ is a monotone decreasing function of t, and it behaves as $2\tau/\alpha t$ for large t. Also p_{a22} will behave in a similar manner because it is bounded by p_{e22} by means of Eq. (32).

The observation time $t_o(\lambda)$ is defined as the time required for the variance of the estimate to become smaller than a specified bound λ. This is immediately computed from Eq. (63) as

$$t_o(\lambda) = \frac{2\tau}{\alpha\lambda}\left[1 - \frac{\lambda}{p_{e22}(0)}\right]. \tag{64}$$

Since $p_{e22}(0)$ usually assumes a much larger value than λ, $t_o(\lambda)$ is simply reduced to

$$t_o(\lambda) \approx \frac{2\tau}{\alpha\lambda} = \frac{1}{\lambda}\frac{N_o}{2A^2}. \tag{65}$$

Therefore, the observation time $t_o(\lambda)$ is inversely proportional to λ, and its coefficient is determined by the signal-to-noise ratio of the input. It is neither dependent on the system gain b nor on the estimation period T. This is a remarkable result, and it is valid for

$$\frac{1}{B_{w_1}} << T << \frac{1}{b}.$$

(B_{w_1} is the bandwidth of noise.)

We now make a comparison of this result against the case without the digital filter. The steady-state variance λ_o of the phase-locked loop having the configuration of Fig. 7 is given by

$$\lambda_o = \frac{b}{4A^2}N_o = \frac{1}{\alpha}. \tag{66}$$

On the other hand, 4τ ($\tau = 1/b$ is the time constant of the loop) may be sufficient for the transient to vanish, so that it is selected as the observation time $t'_o(\lambda)$ without the digital filter. Then, using the relation of Eq. (66),

$$t'_o(\lambda_o) = \frac{4}{b} = \frac{1}{\lambda_o}\frac{N_o}{A^2}. \tag{67}$$

Compared to Eq. (65)

$$t_o(\lambda_o) = \frac{1}{2}t'_o(\lambda_o). \tag{68}$$

Thus, the observation time is reduced, at least to half, by the application of the digital filter for $\lambda = \lambda_o$. Since $t_o(\lambda)$ is the decaying time for p_{e22}, the decaying time of the actual variance p_{a22}, which is bounded by p_{e22}, must be shorter than that in most cases, as is observed in the experimental simulation at the end of this article.

Moreover, the variance of the estimate with the digital filter is monotonically decreasing with time, thus assuring the continuous improvement of the estimate, while the variance remains constant after $t'_o(\lambda_o)$ for the system without the digital filter, and no further improvement is expected on the estimate beyond that point. This is also observed on the experimental result.

Another view of the advantage gained by this filter can be seen in the frequency domain. The bandwidth $B_c(t)$ of the combined system is a function of t; viz.

$$B_c(t) = B_L \cdot \frac{2\tau}{t} \qquad \text{for } t >> 4\tau\frac{\lambda_o}{p_{e22}(0)}. \tag{69}$$

Thus, the combined system behaves as a time varying filter whose bandwidth decreases inversely proportional to time for sufficiently large t, and at time $t = 4\tau$, B_c is already one half of B_L.

5. Experimental Simulation

The entire system including a model of the phase-locked loop, the digital filter, signal and noise generators is simulated on the digital computer. Several experiments have been conducted varying the initial settings of $P_c(0)$ and $x^*_a(0)$ matrices.

The simplest form of the non-negative definite $E_{ea}(0)$ matrix is a diagonal one having

$$e_{ea,11}(0) \geq 0 \tag{70}$$

$$e_{ea,22}(0) > 0. \tag{71}$$

Based on the above relations, diagonal components of $P_e(0)$ matrix are selected so as to satisfy the following condition

$$p_{e11}(0) \gg \sigma_\theta^2 \quad (72)$$

$$p_{e22}(0) > p_{a22}(0) = Z^2 \quad (73)$$

when the initial estimates of states are set at zero,

$$x^*_{1a}(0) = 0 \quad (74)$$

$$x^*_{2a}(0) = 0. \quad (75)$$

Although Z^2 (Z is the amplitude of the input step) is an unknown quantity, we assume that a certain upper bound $\max(Z^2)$ is known. Then $p_{e22}(0)$ is determined at some value no smaller than $\max(Z^2)$.

$$p_{e22}(0) \gg \max(Z^2). \quad (76)$$

Since the effect of the initial value $p_{e22}(0)$ vanishes very quickly, as observed from the solution of the differential equation in Eq. (63), the choice is rather arbitrary as long as it satisfies the above condition.

The rest of the components are assigned as

$$p_{e12}(0) = p_{e21}(0) = 0 \ . \quad (77)$$

In Fig. 8(a), $x_1(k)$ $[= y(k)]$, $x_2(k)$ $(= Z)$ and $x^*_{2a}(k)$ (non-optimum estimate) are plotted. The chained line indicates the range of the calculated variance $p_{e22}(k)$. The experimental results prove that the ensemble of $x^*_{2a}(k)$ remains inside this range verifying the conclusion of Theorem 1. In this particular example, it is observed that $x^*(k)$ enters into the range of $\lambda_o = 0.0025$, the steady-state variance of the phase-locked loop of this example, at $t = 1.7\tau$, and remains within that range afterward, further approaching to the actual input value.

Fig. 8(b) and(c) are the cases when the loop gain b is changed to some other values. Observe that $t_o(\lambda)$ maintains almost the same value irrespective of the variation of the loop gain, and hence the time constant of the loop.

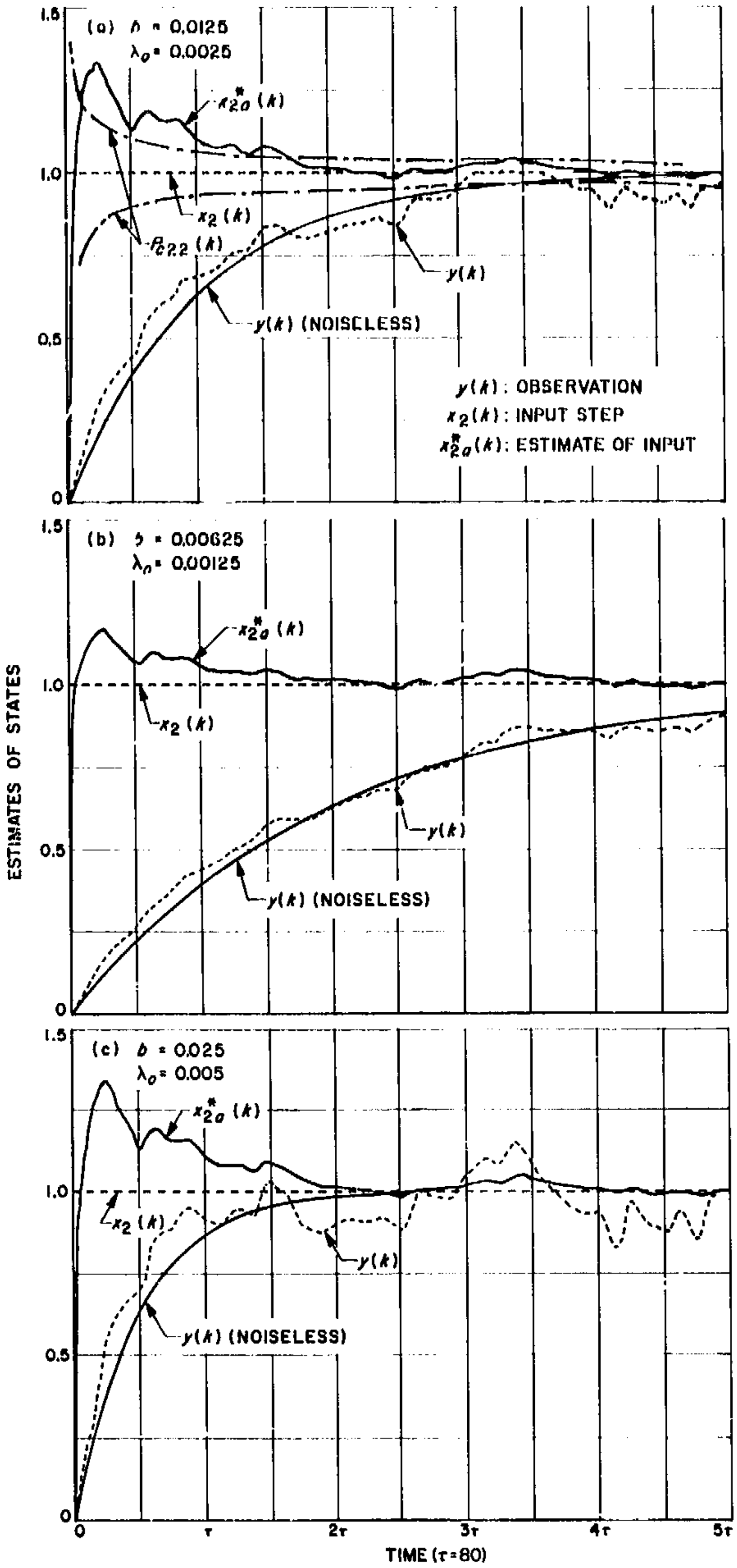


Fig. 8. Experimental simulation

The technique developed in this article can be extended to the case when the signal and noise power are unknown, or when they are varying with time. In that case, it becomes a problem of parameter estimation (Ref. 11), and an optimum estimate may not be obtained. However, under certain conditions, a quasi-optimum estimate can be expected.

Although we have treated the signal (phase) input as a state in this analysis, it is possible to describe the input contribution in an explicit form in Eq. (1) (Ref. 12). Especially when the input to the phase-locked loop takes a form of frequency step or some other form, such explicit description of the input contribution may simplify the treatment of the problem, though it is immaterial for the step input.

F. Cross-Correlation by Computer

P. Schottler

A received signal plus noise (S + N) waveform is to be cross-correlated with a family of locally generated waveforms, the product to be averaged over a time interval T_s. Each of the local waveforms consists of an interval of square wave of length T_s; outside that interval, the waveform is identically zero. The square wave has period $2T$, and T_s satisfies

$$T_s = 2mT$$

$$m = 2^n; \qquad n = 0, 1, 2, \cdots, 7. \qquad (1)$$

The members of the local family of waveforms are identical, except that they are displaced in time relative to one another (Fig. 9).

The cross-correlations are to be performed digitally on a digital computer. The S + N waveform is represented by a sequence of real numbers, which is the output of an analog-to-digital converter which is sampling the received waveform at a constant rate of R samples/second. In a length of time T_s, the sequence generated will contain $R \cdot T_s$ terms.

The samples from the analog-to-digital converter are not read directly into the computer. Rather, the situation is as depicted in Fig. 10. Successive samples are algebraically added into an accumulator during an interval of

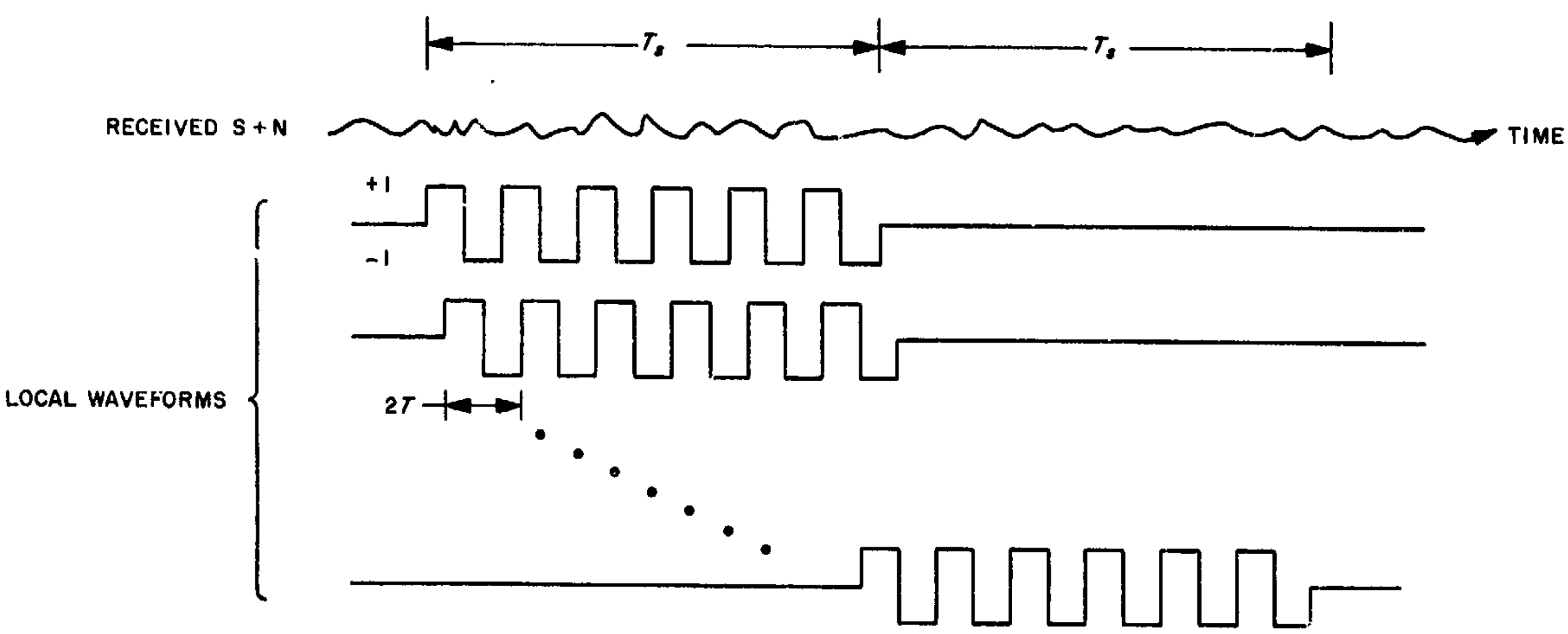


Fig. 9. Family of local waveforms

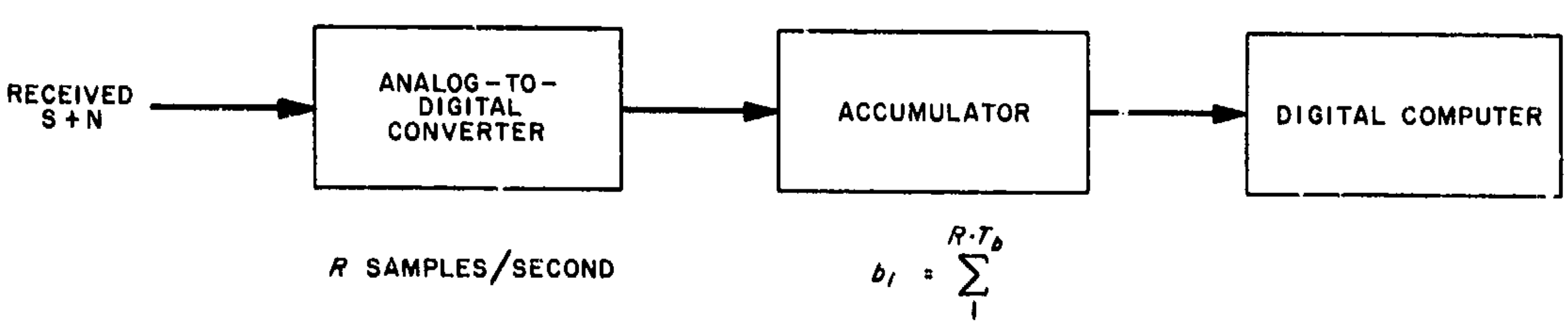


Fig. 10. Computer and input equipment

time of, say, T_b seconds. When $R \cdot T_b$ samples have been accumulated, the resulting sum is read into the computer, the accumulator is reset to zero, and the process then repeats.

To be more specific about the interval of time T_b, consider Fig. 11 which shows 1 cycle of square wave from one of the locally generated waveforms. Divide the half-period T into Q equal intervals, each of time duration

$$T_b = \frac{T}{Q} \tag{2}$$

seconds. In the analysis which follows, Q will be maintained as a variable. Once a value for Q is chosen, the value of T_b is fixed. Also, it is assumed that the family of locally generated waveforms contains $2mQ$ members, spaced T_b seconds apart. Thus, for each interval T_s of received S + N waveform, $2mQ$ cross-correlations must be performed.

In the next part of this article, three procedures are discussed for computing the required cross-correlations between the received S + N and the family of local waveforms. In Part 2, on-line processing is discussed.

1. Correlation Procedures

It is first necessary to introduce some terminology so that the procedures can be discussed in terms of computer operations.

Let $a^{\dagger} = \{a_1, a_2, \cdots\} = \{a_i\}$ be a sequence of real numbers (the samples). Define a sequence $b^{\dagger} = \{b_1, b_2, \cdots\}$ by

$$b_i = \sum_{j=(i-1)N+1}^{iN} a_j , \tag{3}$$

where $N = T_b \cdot R$, a positive integer (inputs to the computer from the accumulator).

Let b be the finite sequence from $b^{\dagger}$, containing $4mQ$ terms

$$b = \{b_1, b_2, \cdots, b_{2mQ}, b_{2mQ+1}, \cdots, b_{4mQ}\} \tag{4}$$

(corresponds to two successive intervals of length T_s of received S + N).

Define a family of finite sequences, containing $2mQ$ members, each member having $4mQ$ terms (the family of local waveforms)

$$y(0) = \{\overbrace{\underbrace{+1, +1, \cdots, +1}_{Q}, \underbrace{-1, -1, \cdots, -1}_{Q}, \cdots, \underbrace{-1, -1, \cdots}_{Q}}^{2mQ \text{ terms}}, \overbrace{1, 0, 0, \cdots, 0}^{2mQ \text{ terms}}\}$$

$$y(1) = \{0, \underbrace{+1, +1, \cdots, +1}_{Q}, \underbrace{-1, -1, \cdots, -1}_{Q}, \cdots, \underbrace{-1, -1, \cdots, -1}_{Q}, \overbrace{0, 0, \cdots, 0}^{2mQ-1 \text{ terms}}\}$$

$$\vdots$$

$$y(2mQ-1) = \{\underbrace{0, 0, \cdots, 0}_{2mQ-1}, \underbrace{+1, +1, \cdots, +1}_{Q}, \underbrace{-1, -1, \cdots, -1}_{Q}, \cdots, \underbrace{-1, -1, \cdots, -1}_{Q}, 0\}. \tag{5}$$

Finally, define a term-by-term product for the sequences b and $y(j)$

$$c_i(j) = b_i \cdot y_i(j) \tag{6}$$

and a weight function

$$S_j = \sum_{i=1}^{4mQ} c_i(j) \tag{7}$$

for the product sequences $c(j)$. The weight functions, $\{S_j\}_{j=0}^{2mQ-1}$, are the required cross-correlations.

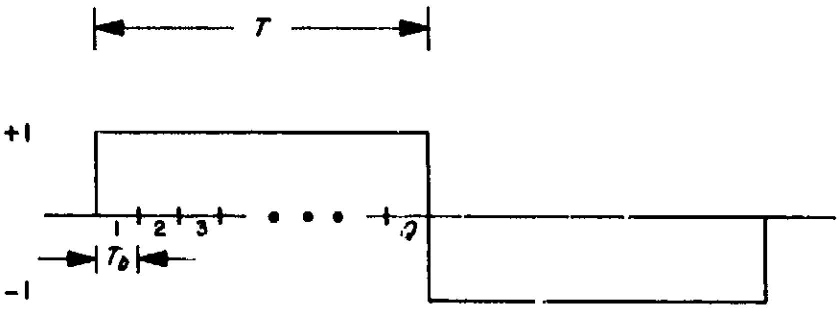


Fig. 11. Subdivision of T

The three correlation procedures to be discussed will be compared on the basis of the number of operations (additions or subtractions) which the computer must perform to obtain the $\{S_M\}_{M=0}^{2mQ-1}$. Let J_i be the total number of operations required to compute all S_M using the ith procedure, and write

$$J_i = J_i(0) + (2mQ - 1)J_i(M), \tag{8}$$

where $J_i(0)$ is the number of operations required to compute S_0, given $b_1, b_2, \cdots, b_{2mQ}$ and $J_i(M)$ is the number of operations required to compute S_M, given $b_1, \cdots, b_{2mQ+M}, S_0, S_1, \cdots, S_{M-1}$.

a. Procedure 1. Using Eqs. (4) through (7), the first two weight functions are

$$\begin{aligned} S_0 = {} & [b_1 + b_2 + \cdots + b_Q] - [b_{Q+1} + b_{Q+2} + \cdots \\ & + b_{2Q}] + \cdots + [b_{2Q(m-1)+1} + \cdots + b_{Q(2m-1)}] \\ & - [b_{Q(2m-1)+1} + \cdots + b_{2mQ}] \\ = {} & \sum_{k=1}^{2m} (-1)^{k+1} \sum_{l=Q(k-1)+1}^{Qk} b_l \end{aligned}$$

and

$$\begin{aligned} S_1 = {} & [b_2 + b_3 + \cdots + b_Q + b_{Q+1}] - [b_{Q+2} + \cdots \\ & + b_{2Q} + b_{2Q+1}] + \cdots + [b_{2Q(m-1)+2} + \cdots \\ & + b_{Q(2m-1)+1}] - [b_{Q(2m-1)+2} + \cdots + b_{2mQ} + b_{2mQ+1}] \\ = {} & \sum_{k=1}^{2m} (-1)^{k+1} \sum_{l=Q(k-1)+2}^{Qk+1} b_l. \end{aligned}$$

In general, the Mth weight function is given by

$$S_M = \sum_{k=1}^{2m} (-1)^{k+1} \sum_{l=Q(k-1)+M+1}^{Qk+M} b_l. \tag{9}$$

Using Eq. (9), the computer must do $2mQ$ additions or subtractions for each S_M. Thus

$$J_1(0) = J_1(M) = 2mQ$$

and

$$J_1 = 2mQ + (2mQ - 1)\,2mQ = (2mQ)^2. \tag{10}$$

b. Procedure 2. The second procedure makes use of an iterative technique given in the following.

Theorem.

$$S_{M+1} = S_M - b_{M+1} + 2\sum_{k=1}^{2m-1} (-1)^{k+1} b_{Qk+M+1} - b_{2mQ+M+1}. \tag{11}$$

Proof. From Eq. (9)

$$\begin{aligned} S_{M+1} = {} & \sum_{k=1}^{2m} (-1)^{k+1} \sum_{l=Q(k-1)+M+2}^{Qk+M+1} b_l \\ = {} & \sum_{k=1}^{2m} (-1)^{k+1} \left[\sum_{l=Q(k-1)+M+1}^{Qk+M} b_l + b_{Qk+M+1} - b_{Q(k-1)+M+1} \right] \\ = {} & \sum_{k=1}^{2m} (-1)^{k+1} \sum_{l=Q(k-1)+M+1}^{Qk+M} b_l + \sum_{k=1}^{2m} (-1)^{k+1} b_{Qk+M+1} \\ & - \sum_{k=1}^{2m} (-1)^{k+1} b_{Q(k-1)+M+1}. \end{aligned}$$

Let $j = k - 1$, then

$$\begin{aligned} S_{M+1} = {} & S_M + \sum_{k=1}^{2m} (-1)^{k+1} b_{Qk+M+1} - \sum_{j=0}^{2m-1} (-1)^{j+2} b_{Qj+M+1} \\ = {} & S_M + \sum_{k=1}^{2m-1} (-1)^{k+1} b_{Qk+M+1} + (-1)^{2m+1} b_{2mQ+M+1} \\ & - (-1) \sum_{j=0}^{2m-1} (-1)^{j+1} b_{Qj+M+1} \\ = {} & S_M + \sum_{k=1}^{2m-1} (-1)^{k+1} b_{Qk+M+1} - b_{2mQ+M+1} \\ & + \sum_{j=1}^{2m-1} (-1)^{j+1} b_{Qj+M+1} - b_{M+1} \\ = {} & S_M - b_{M+1} + 2\sum_{k=1}^{2m-1} (-1)^{k+1} b_{Qk+M+1} - b_{2mQ+M+1}. \end{aligned}$$

Using the iterative technique, the computation of S_{M+1}, given the b_i and S_M, requires $2m + 3$ additions or subtractions. Thus,

$$J_2(M) = 2m + 3. \tag{12}$$

The weight function S_0 must, of course, be computed from Eq. (9), giving

$$J_2(0) = 2mQ. \tag{13}$$

Thus, using Eq. (11), the total number of computer operations required is

$$J_2 = 2mQ + (2mQ - 1)(2m + 3). \tag{14}$$

Comparison of the first two procedures on the basis of the number of computer operations yields

$$\frac{J_2 - J_2(0)}{J_1 - J_1(0)} = \frac{1}{Q}\left(1 + \frac{3}{2m}\right) < 1 \qquad Q \geq 3, \text{ all } m. \tag{15}$$

It follows from Eq. (15) that Procedure 2 should always be used in preference to Procedure 1 if $Q \geq 3$, or if $Q \geq 2$, $m > 1$, etc.

c. Procedure 3. Expand Eq. (9) as follows

$$S_M = \sum_{k=1}^{2m} (-1)^{k+1} \sum_{l=Q(k-1)+M+1}^{Qk+M} b_l$$

$$= \sum_{k=1}^{2m} (-1)^{k+1} [b_{Q(k-1)+M+1} + b_{Q(k-1)+M+2} + \cdots + b_{Qk+M}]. \qquad (16)$$

Now, define the Q sums

(17·1) $\quad S_M\{b_{Q(k-1)+M+1}\} = \sum_{k=1}^{2m} (-1)^{k+1} b_{Q(k-1)+M+1}$

$\cdot$

$\cdot$

$\cdot$

(17·Q) $\quad S_M\{b_{Qk+M}\} = \sum_{k=1}^{2m} (-1)^{k+1} b_{Qk+M}, \qquad (17)$

so that Eq. (16) becomes

$$S_M = S_M\{b_{Q(k-1)+M+1}\} + \cdots + S_M\{b_{Qk+M}\}. \qquad (18)$$

Note that $S_M\{b_{Q(k-1)+M+1}\}$ contains b_i with $i \equiv M+1 \pmod{Q}$, $S_M\{b_{Q(k-1)+M+2}\}$ contains all b_i with $i \equiv M+2 \pmod{Q}$, etc.

Expand S_{M+1} as in Eq. (16)

$$S_{M+1} = \sum_{k=1}^{2m} (-1)^{k+1} \sum_{l=Q(k-1)+M+2}^{Qk+M+1} b_l$$

$$= \sum_{k=1}^{2m} (-1)^{k+1} [b_{Q(k-1)+M+2} + \cdots + b_{Qk+M} + b_{Qk+M+1}]$$

$$= S_{M+1}\{b_{Q(k-1)+M+2}\} + \cdots + S_{M+1}\{b_{Qk+M}\} + S_{M+1}\{b_{Qk+M+1}\} \qquad (19)$$

with

(20·1) $\quad S_{M+1}\{b_{Q(k-1)+M+2}\} = \sum_{k=1}^{2m} (-1)^{k+1} b_{Q(k-1)+M+2}$

$\cdot$

$\cdot$

$\cdot$

(20·Q-1) $\quad S_{M+1}\{b_{Qk+M}\} = \sum_{k=1}^{2m} (-1)^{k+1} b_{Qk+M} \qquad (20)$

(20·Q) $\quad S_{M+1}\{b_{Qk+M+1}\} = \sum_{k=1}^{2m} (-1)^{k+1} b_{Qk+M+1}.$

Comparison of Eq. (20) with (17) reveals that

$$S_{M+1}\{b_{Q(k-1)+M+2}\} = S_M\{b_{Q(k-1)+M+2}\} \qquad (21)$$

$\cdot$

$\cdot$

$\cdot$

$$S_{M+1}\{b_{Qk+M}\} = S_M\{b_{Qk+M}\}$$

In (20·Q), let $k = j-1$, then

$$S_{M+1}\{b_{Qk+M+1}\} = \sum_{j=2}^{2m+1} (-1)^j b_{Q(j-1)+M+1}$$

$$= -\sum_{j=1}^{2m} (-1)^{j+1} b_{Q(j-1)+M+1} + (-1)^{2m+1} \times b_{2mQ+M+1} + b_{M+1}$$

or

$$S_{M+1}\{b_{Qk+M+1}\} = -[S_M\{b_{Q(k-1)+M+1}\} - b_{M+1}] - b_{2mQ+M+1}. \qquad (22)$$

The significance of Eqs. (21) and (22) is clear. If the Q sums $S_M\{b_{Q(k-1)+M+1}\}, \cdots, S_M\{b_{Qk+M}\}$ are maintained in the computer, together with the b_i and the S_M, then the weight function S_{M+1} can be computed by first computing $S_{M+1}\{b_{Qk+M+1}\}$, using Eq. (22), and then solving

$$S_{M+1} = S_M - S_M\{b_{Q(k-1)+M+1}\} + S_{M+1}\{b_{Qk+M+1}\}. \qquad (23)$$

The computation of $S_{M+1}\{b_{Qk+M+1}\}$ via Eq. (22) requires three operations, and (23) also requires three operations, hence

$$J_3(m) = 3 + 3 = 6 \qquad (24)$$

independent of both m and Q.

In order to compute S_0 using Procedure 3, it is necessary first to compute the Q sums $S_0\{b_{Q(k-1)+1}\}, \cdots, S_0\{b_{Qk}\}$. Each of these Q sums requires $2m$ additions or subtractions. Then these Q sums must be added together to obtain S_0. Thus

$$J_3(0) = 2mQ + Q \qquad (25)$$

giving

$$J_3 = (2m+1)Q + (2mQ-1)6. \qquad (26)$$

Comparison of Eq. (26) with (14) gives

$$\frac{J_3 - J_3(0)}{J_2 - J_2(0)} = \frac{6}{2m+3} < 1 \qquad m > 1. \tag{27}$$

2. On-Line Processing

In each of the three procedures discussed in Part *1*, a new S_M is computed following each b_i input to the computer (after the initial $2mQ$ b_i have been received). It follows that the time required to compute S_M must be less than $T_b = T/Q$, if processing is to be done "on-line."

In an on-line situation, it will be desirable to have the special purpose equipment furnish the computer with S_0 as well as the b_i if the calculation of S_0 requires a significantly greater number of operations than does the calculation of $S_M, M \neq 0$. This is not the case for Procedure 1, since, from Eq. (10), all S_M require the same number of operations. It is the case for Procedures 2 and 3, however, for most values of m and Q. Here it is undesirable for the computation rate to be set by S_0, since that sum is calculated only once every T_s seconds.

The additional special purpose equipment required to furnish S_0 using Procedure 3 is shown in Fig. 12. The b_i are formed as before and read into the computer. In addition, the b_i are multiplied by $y(0)$ and added into a second accumulator which sums over T_s seconds. After T_s seconds, this accumulator will hold

$$\sum_{i=1}^{2mQ} b_i \cdot y_i(0) = S_0.$$

Finally, the $b_i \cdot y_i(0)$ outputs from the multiplier are read into the computer, so that the Q sums

$$S_0\{b_{Q(k-1)+1}\}, \cdots, S_0\{b_{Qk}\},$$

can be formed as the b_i appear. Thus, at the end of the initial T_s second interval, the computer contains S_0, the Q sums $S_0\{b_{Q(k-1)+1}\}$, etc., and the first $2mQ$ b_i. Clearly, the special purpose equipment in Fig. 12 is necessary and sufficient to furnish S_0 for Procedures 1 and 2 as well, although in these two cases the $b_i \cdot y_i(0)$ need not be read into the computer.

Sample programs for computing $S_M, M \neq 0$, for each of the three procedures discussed in Part *1*, are sketched below. No claim to completeness is made for these programs as, for example, indexing operations are not included. The timing for each operation is expressed in units of T_M, the basic memory cycle time, assuming a "typical" digital computer.

a. Procedure 1.

$$S_M = \sum_{k=1}^{2m} (-1)^{k+1} \sum_{l=Q(k-1)+M+1}^{Qk+M} b_l.$$

			Timing
$k = 1$	Load	b_{M+1}	2
	Add	b_{M+2}	2
	.		
	.		
	.		
	Add	b_{Q+M}	2
	Store	α_1	3
			$3 + 2Q$
$k = 2$	Load	b_{Q+M+1}	
	Add		
	.		$3 + 2Q$
	.		
	.		
.			
.	Store	α_2	
$k = 2m$	.		
	.		$3 + 2Q$
	.		
	Store α_{2m}		
	Load α_1		2
	Sub α_2		2
	Add α_3		
	.		
	.		
	.		
	Sub α_{2m}		2
	Store		3
			$2(2m) + 3$

$$T_1 = [2m(3 + 2Q) + 2(2m) + 3]\, T_M$$

$$T_1 = [4mQ + 10m + 3]\, T_M \tag{28}$$

b. Procedure 2.

$$S_{M+1} = S_M - b_{M+1} + 2 \sum_{k=1}^{2m-1} (-1)^{k+1} b_{Qk+M+1} - b_{2mQ+M+1}.$$

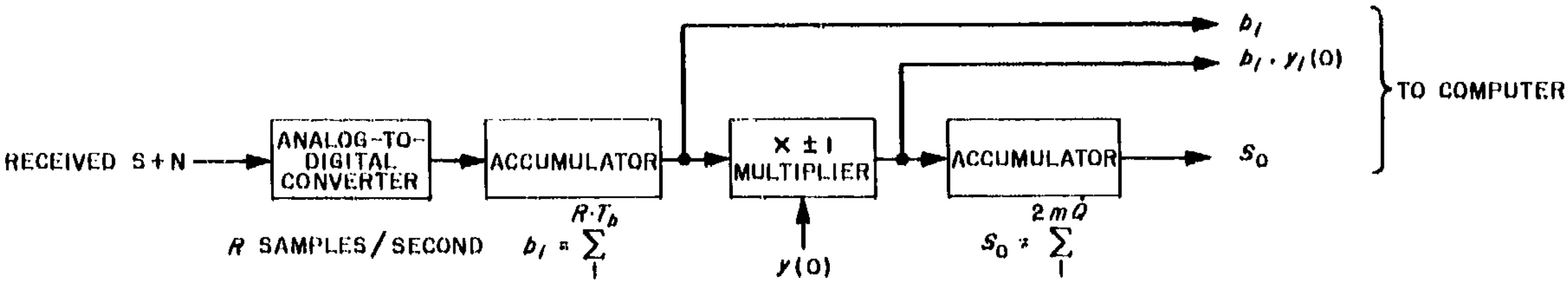


Fig. 12. On-line input equipment

		Timing
Load	b_{Q+M+1}	2
Sub	b_{2Q+M+1}	2
Add		2
.		
.		
.		
Add		2
Left shift		2
Store	α_1	3
		$2(2m-1)+2+3$

Load	S_M	2
Sub	b_{M+1}	2
Add	α_1	2
Sub	$b_{2mQ+M+1}$	2
Store		3
		11

$$T_2 = [4m+3+11]\,T_m = [4m+14]\,T_M. \qquad (29)$$

c. Procedure 3.

$$S_{M+1}\{b_{Qk+M+1}\} = S_M\{b_{Q(k-1)+M+1}\} + b_{M+1} - b_{2mQ+M+1}$$

$$S_{M+1} = S_M - S_M\{b_{Q(k-1)+M+1}\} + S_{M+1}\{b_{Qk+M+1}\}.$$

		Timing
Load	b_{M+1}	2
Sub	$S_M\{b_{Q(k-1)+M+1}\}$	2
Sub	$b_{2mQ+M+1}$	2
Store	α_1	3
		9
Load	S_M	2
Sub	$S_M\{b_{Q(k-1)+M+1}\}$	2
Add	α_1	2
Store		3
		9

$$T_3 = [9+9]\,T_M = 18\,T_M. \qquad (30)$$

Fig. 13 shows a graph of T_1/T_M, T_2/T_M, and T_3/T_M as a function of m, with Q a parameter [Eqs. (28) through (30)].

It is clear, from Fig. 13, that Procedure 3 is never slower than either Procedure 1 or 2, and for most values of m it is significantly faster than both Procedures 1 and 2. If Procedure 3 is used on-line, then

$$T_3 = 18\,T_M < T_b = \frac{T}{Q}, \qquad (31)$$

from which it follows that

$$Q < \frac{T}{18\,T_M}. \qquad (32)$$

Assuming that $T = 200\ \mu$sec, $T_M = 1.75\ \mu$sec, then Eq. (32) gives

$$Q < 6.3$$

or, since Q must be an integer which divides T,

$$Q = 1, 2, 4, \text{or } 5.$$

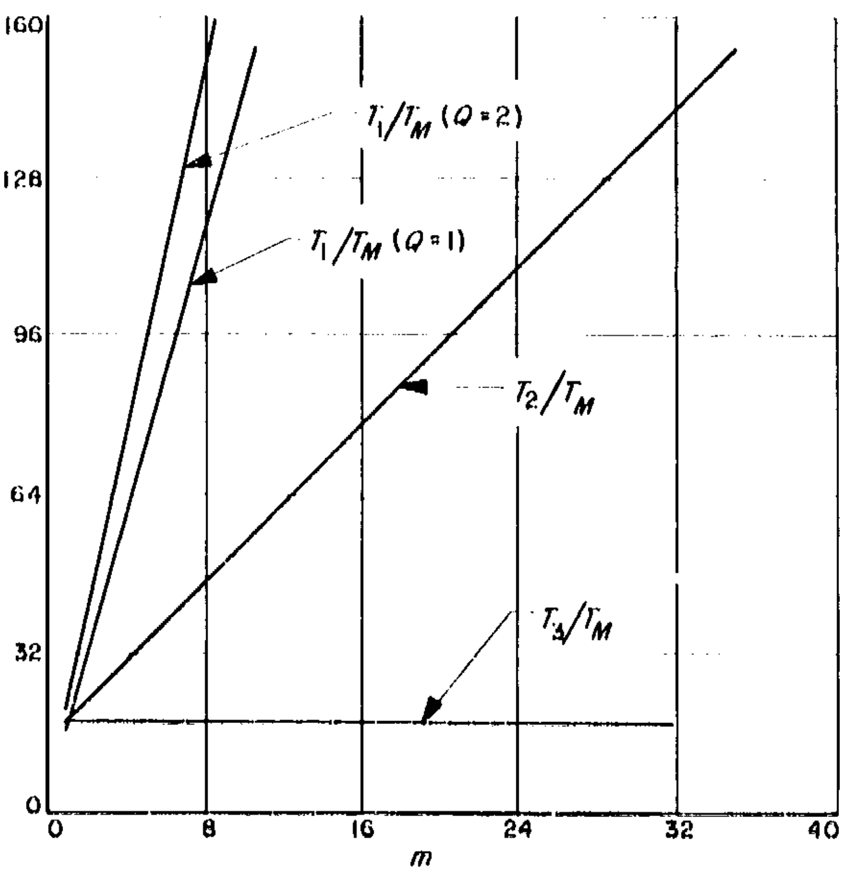


Fig. 13. Execution times for sample programs

G. Performance Parameters for Sequential Decoding

I. Jacobs

The application of sequential decoding to the communication of discrete data over an additive Gaussian noise channel is evaluated in terms of the energy per bit required for reliable one-way operation. The storage requirements of several receiver quantization schemes are detailed, and attention is focused on the problem of buffer overflow. A phase-coherent system utilizing coded binary antipodal signals and a detector with 3-bit quantization is found to combine nearly optimum efficiency with reasonable storage requirements. The potential advantage over a system using 32 biorthogonal signals (with no further coding) is 3.4 to 4 db at a word error probability of 10^{-5}.

1. The Need for Coding

The use of an alphabet of M equal-energy orthogonal signals to transmit $\log_2 M$ bits of information over an additive white Gaussian noise channel is essentially optimum for both coherent and incoherent operation if (1) sufficient bandwidth is available and (2) a maximum likelihood receiver is used. It follows that orthogonal signals can be used to transmit at rates arbitrarily close to the infinite bandwidth channel capacity with arbitrarily low error probability. An equivalent statement is that as long as the energy received per bit of information E_b exceeds $E_{b,min} = N_0 \ln 2$, where N_0 is the received noise power in a bandwidth of 1 cps, the word error probability P_E (and hence the bit error probability $\approx P_E/2$) tends to zero as M increases. A useful bound (Ref. 13) is

$$P_E < \begin{cases} M^{-[(E_b/E_{b,min})^{1/2} - 1]^2}, & E_b/E_{b,min} \leq 4 \\ M^{-[(E_b/2E_{b,min}) - 1]}, & E_b/E_{b,min} > 4. \end{cases} \tag{1}$$

Eq. (1) serves to illustrate the difficulty encountered whenever both efficient operation, E_b close to $E_{b,min}$, and low error probability are simultaneously demanded; the required value of M becomes excessively large as indicated in Table 1. Recognizing that at most $2TW$ orthogonal signals can be transmitted in time T and bandwidth W, we observe that to transmit at a rate of 3 bits/sec with $M = 10^{11}$ requires an information bandwidth of about 4 Gc, while $M = 10^{20}$ requires 10^{18} Gc. Clearly, very low error, very efficient communication with orthogonal signals is precluded by bandwidth considerations. Furthermore, maximum likelihood receivers cannot be built with such large M, so that the performance indicated in Table 1 cannot actually be attained.

Table 1. Required size of orthogonal signal alphabet, *M*

Bound	$E_b/E_{b,min}$	
	At 3 db	At 4.5 db
10^{-3}	$10^{12.5}$	$10^{8.5}$
10^{-5}	$10^{29.2}$	$10^{16.9}$
10^{-7}	10^{41}	$10^{18.3}$

We conclude that to simultaneously achieve low error probability and high efficiency requires a large nonorthogonal signal set constructed in such a way that a near optimum receiver can be realized. An attractive possibility, which we next consider in detail, incorporates convolutional encoding and sequential decoding.

2. System Description

A rough block diagram is presented in Fig. 14. Two modulators are treated. One, for use on a channel with coherent phase reference, generates a sequence of binary antipodal waveforms, each of energy E, under direction of the coder. These waveforms are referred to as channel symbols. The detector consists of a filter matched to the

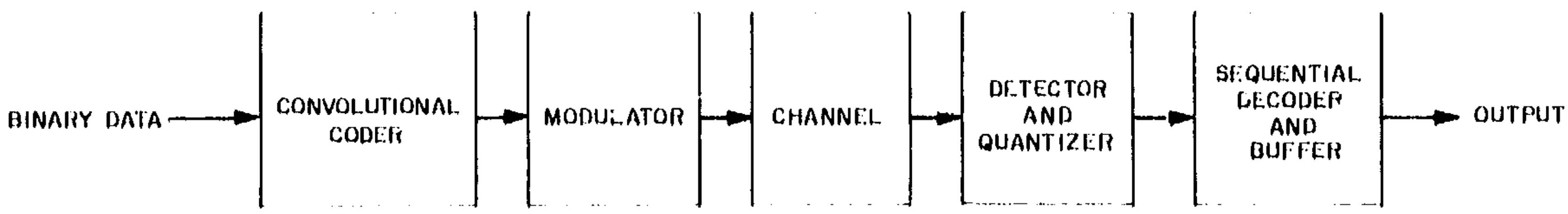


Fig. 14. System block diagram

individual waveforms followed by a time sampler and one of the three quantizers of Fig. 15. For each R information bits supplied to the coder, one channel symbol is transmitted and $\log_2 Q$ bits generated at the detector output and loaded into the decoder (a digital machine). The energy per bit is, therefore

$$E_b = \frac{E}{R}, \tag{2}$$

while the number of bits produced at the detector output (which must be stored in the decoder) per information bit, denoted β, is

$$\beta = \frac{1}{R}\log_2 Q. \tag{3}$$

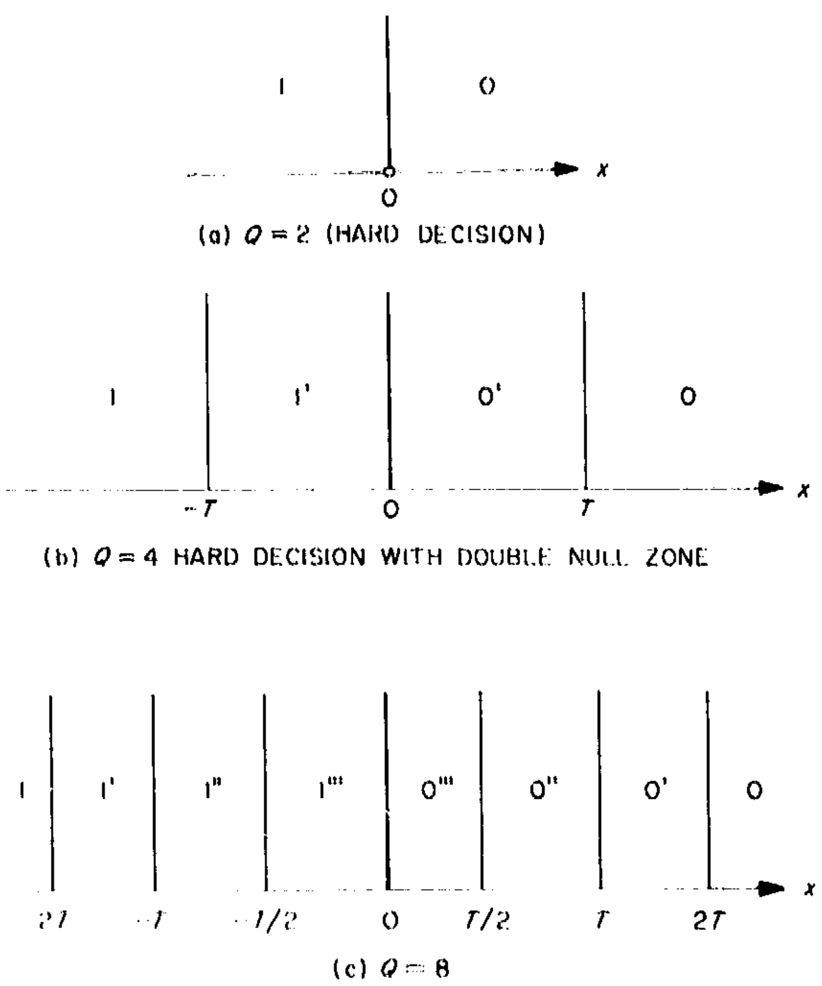


Fig. 15. Detector quantization schemes for system using binary antipodal signal; x denotes the time-sampled matched filter output

The second modulator to be treated, suitable for an incoherent channel, generates a sequence of waveforms (channel symbols) chosen from an alphabet of 32 orthogonal signals, each of energy E. If R again denotes the information rate in bits per channel symbol, Eq. (2) remains valid. For this case, we assume that the detector consists of a bank of matched filters, square-law envelope detectors, and samplers. For each channel symbol received, the detector produces as output an ordered list of the L most likely waveforms. Since 5 bits are sufficient to identify a letter from a 32-letter alphabet, the value of Q for use in Eq. (3) is

$$\log_2 Q = 5L. \tag{4}$$

The cascade of modulator, channel, and detector which connects the coder output to the decoder input is usefully modeled as a memoryless discrete channel with A input letters ($A = 2$ for the coherent system and $A = 32$ for the incoherent system) and Q output letters. Each transmission of a channel symbol over the original channel corresponds to one use of the discrete channel. A discrete channel is specified by a transition probability matrix $\{q_{ij}\}$ where q_{ij} is the probability that j is received when i is transmitted. The matrix $\{q_{ij}\}$ determines a parameter, R_0, the computational cutoff rate, of central importance to sequential decoding (Refs. 13–15). For the channels considered,

$$R_0 = -\log \sum_{j=1}^{Q}\left[\sum_{i=1}^{A}\frac{1}{A}(q_{ij})^{1/2}\right]^2 \tag{5}$$

For the infinite bandwidth Gaussian channel and maximum likelihood reception, R_0 equals one-half channel capacity. For a finite bandwidth signal set and nonmaximum likelihood reception, R_0 is less than one-half of the infinite bandwidth channel capacity. Thus, if $R \leq R_0$,

$$E_b \geq \frac{E}{R_0}, \tag{6}$$

and

$$10\log_{10}\left(\frac{E_b}{E_{b,min}}\right) \geq 3\,\text{db}. \tag{7}$$

3. System Efficiency

In this part of this article, we determine the efficiency of the coherent and incoherent systems assuming operation at $R = R_0$ bits per channel use. In the next part, we argue that in an operational system R must be less than R_0, and determine the resulting effect on efficiency.

For each value of signal energy E and choice of T in Fig. 15, it is straightforward to determine the transition probabilities $\{q_{ij}\}$ and thence R_0, $E_b/E_{b,min}$, and β. A digital computer was used to select T to minimize $E_b/E_{b,min}$. The result is plotted in Fig. 16 versus β for both the incoherent and coherent systems.

In the coherent system, decreasing E (the energy in each antipodal signal) results in a smaller decrease in R_0, thus yielding a decrease in $E_b/E_{b,min}$ while increasing β (and bandwidth). For fixed E, increasing Q also decreases $E_b/E_{b,min}$ and increases β. It is important to note that with $\beta = 15$ storage bits/information bit and $Q = 8$, $E_b/E_{b,min}$ is only ½ db above the minimum specified by Eq. (7). Thus, binary antipodal signaling and 3-bit detector quantization is very efficient. (It is possible that transmission of letters from a biorthogonal signal alphabet may permit β to be decreased without increase in $E_b/E_{b,min}$, although curves in Ref. 16 seem to indicate the opposite.)

In the incoherent system, $E_b/E_{b,min}$ does not continue to decrease as the energy in each signal decreases. Square-law envelope detectors become inefficient when the signal energy is small relative to the noise. This weak-signal suppression effect causes the minimums observed in the curves of Fig. 15. The incoherent system is thus considerably less efficient than the coherent system. Even if maximum likelihood incoherent reception is used, it can be shown (Ref. 14) that for this incoherent system $E_b/E_{b,min}$ always exceeds 5.2 db. The use of a coherent phase reference appears to be especially desirable with sequentially coded systems.

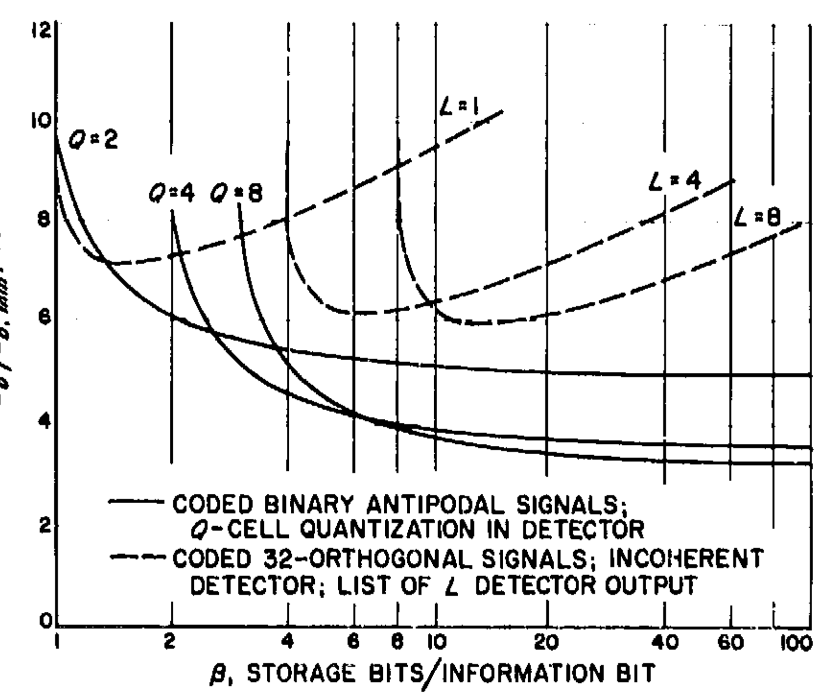


Fig. 16. Energy requirements versus storage requirements for two systems utilizing sequential decoding at rate $R = R_0$

4. The Computation Problem

In a properly designed sequential decoder utilizing a constraint length of 60 to 100 bits, the probability of error is negligibly small for rates R less than R_0 (Ref. 17). The variability of computation is then the limiting factor. It is shown[6] that the average number of computations per decoded bit C obeys a Parèto law:

$$P[C > L] \approx L^{-\alpha}, \qquad L > 1. \tag{8}$$

For integer values, the Parèto exponent α is determined by the construction of Fig. 17. It is hypothesized that this construction is valid for all α. For operating conditions of interest, the sphere-packing exponent is very nearly that of the very noisy channel (Ref. 13).

$$\text{Exponent} = [(2R_0)^{1/2} - (R)^{1/2}]^2. \tag{9}$$

It follows that

$$\alpha = 2\frac{R_0}{R} - 1. \tag{10}$$

[6] "The Computation Problem With Sequential Decoding," by J. E. Savage, Ph.D. Thesis, Department of Electrical Engineering, Cambridge, Mass., February 1965.

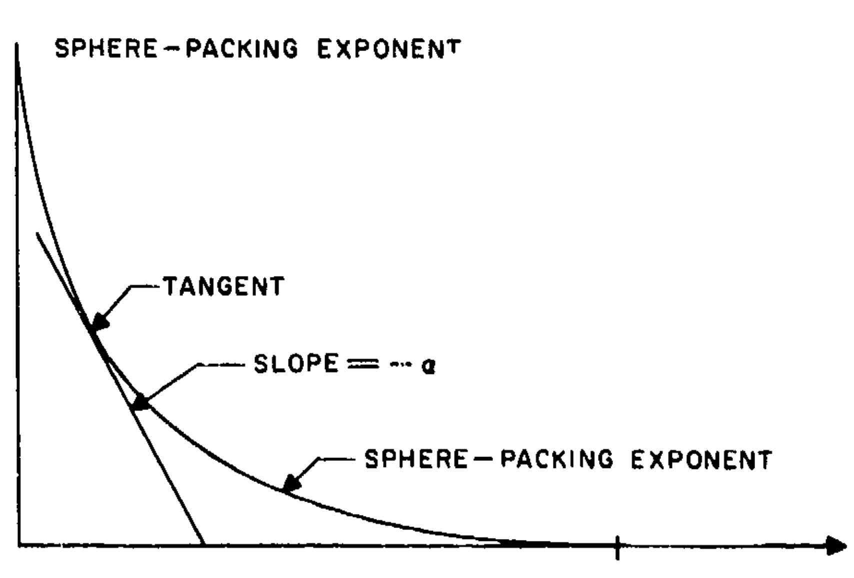


Fig. 17 Determination of α as a function of R

The Parèto law of Eq. (8) is notable for the large probability of obtaining very large values of C. To accommodate the system to this extreme variability in computations per bit, a large buffer must be provided with the decoder. (The importance of operating with small β is now clear.) Indeed, in implementing the Fano algorithm, the buffer is made an integral part of the decoder (Ref. 15). Unfortunately, finite buffers have non-zero probability of overflow. Whenever overflow occurs, data is lost and a resynchronization cycle is necessary.

One requirement which must be satisfied to keep the probability of buffer overflow small is that the expected value of C be finite. For the Parèto law of Eq. (8), the expectation is finite only if $\alpha > 1$, hence, by Eq. (10), only if $R < R_0$. For example, if $R = 0.89\,R_0$, resulting in a ½-db increase in $E_b/E_{b,min}$, the expected number of computations per decoded bit is just under 5. It is necessary that the decoder have sufficient computational speed to perform at least 5 times the expected number of computations, or 25, in the time interval separating information bits. Using this factor of 25, we can estimate the probability of buffer overflow. Assume that the buffer can store $\lambda = 4000$ words, each containing β bits. Assume further that the buffer is empty at the time the machine begins to process a particular bit. If the number of computations on this bit exceeds 25λ, the buffer must fill and overflow. From Eq. (8), the probability of such an occurrence with $R = 0.88\,R_0$ is

$$P\,[C > 25\lambda] \approx (25\lambda)^{-1.24} = 10^{-6.2}, \qquad \lambda = 4000. \tag{11}$$

Observe that an order of magnitude increase in either the speed of computation or the size of the buffer causes little more than an order of magnitude decrease in the probability of overflow.

Although further investigation is required, computer simulations (Ref. 18) suggest that the probability of overflow in N consecutive decodings is approximately N times the probability of overflow per decoding, assuming an initially empty buffer. If so, the probability of overflow in 1000 consecutive decodings when Eq. (11) is valid approaches 10^{-3} and completely dominates the probability of error. The probability of overflow in 1000 decodings can be reduced to 10^{-5}, again with $\lambda = 4000$, by setting $R = 0.77\,R_0$. This setting results in an increase in $E_b/E_{b,min}$ of 1.15 db over the values given in Fig. 16.

Whenever feedback from receiver to transmitter is not available, the finite probability of buffer overflow necessitates a provision for periodic resynchronization of the decoder. This may be accomplished by interrupting the data system every 1000 bits and transmitting at most 100 data bits to reset the decoder and to provide a synchronizing word. This decreases the information rate by a factor of 0.9, implying an increase in $E_b/E_{b,min}$ of ½ db.

It thus appears that a system utilizing coded binary antipodal signaling can operate on a coherent additive Gaussian noise channel with negligible error probability and a word erasure probability of 10^{-3} with $E_b/E_{b,min} = 4.5$ db and a buffer containing 4000 17-bit words. The erasure probability can be decreased to 10^{-5} by increasing $E_b/E_{b,min}$ to 5.1 db and the word size to 20 bits.

5. Comparison with Biorthogonal Signaling

The efficiency of a system utilizing biorthogonal signaling on a coherent channel for several word error probabilities and alphabet sizes is given in Table 2 (Ref. 2, pp. 3–14). Thus, if word error probabilities of 10^{-3} can be tolerated, sequential decoding has little to offer. It appears likely, however, that systems which might use sequential decoding will certainly perform considerable data reduction prior to data transmission. A word error probability of 10^{-5} or lower would then be necessary. A system utilizing sequential decoding potentially operates 3.4 to 4 db more efficiently than a system with an alphabet of 32 biorthogonal signals; it operates 1.4 to 2 db more efficiently than a system with an alphabet of 1024 biorthogonal signals. The degradation caused by the use of a phase-locked loop to provide the phase reference will be reported in a future *SPS*.

Table 2. $E_b/E_{b,min}$ for biorthogonal signaling

Word error probability	Alphabet size	
	32	1024
10^{-3}	6.7 db	5.0 db
10^{-5}	8.6 db	6.5 db

References

1. Viterbi, A. J., "Systematic Coding for the Continuous Gaussian Channel," Ph.D. Thesis, University of Southern California, August 1962.

2. Viterbi, A. J., "On Coded Phase-Coherent Communications" *IRE Transactions on Space Electronics and Telemetry*, Vol. SET-7, March 1961.

3. Davenport, W. B., Jr., and Root, W. L., *Random Signals and Noise*, McGraw-Hill Book Co. Inc., New York, 1958.

4. Arthurs, E., and Dym, H., "On the Optimum Detection of Digital Signals in the Presence of White Gaussian Noise," *IRE Transactions on Communications Systems*, December 1962.

5. Jaffe, R. M., and Rechtin, R., "Design and Performance of Phase-Lock Circuits," *Transactions of the IRE*, Vol. IT-1, pp. 66–76, March 1955.

6. Luke, Y. L., *Integrals of Bessell Functions*, McGraw-Hill Book Company, Inc., New York, 1962.

7. Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," *Transactions of ASME, Series D, Journal of Basic Engineering*, Vol. 82, pp. 35–45, March 1960.

8. Kalman, R. E., and Bucy, R. S., "New Results in Linear Filtering and Prediction Theory," *Transactions of ASME, Series D, Journal of Basic Engineering*, Vol. 83, pp. 95–107, March 1961.

9. Soong, T. T., "On A Priori Statistics in Minimum Variance Estimation Problems," Preprints of 1964 Joint Automatic Control Conference at Stanford University, pp. 338–391, June 1964.

10. Ho, Y. C., and Lee, R. C. K., "A Bayesian Approach to Problems in Stochastic Estimation and Control," *IEEE Transactions on Automatic Control*, Vol. 89, No. 4, pp. 333–339, October 1964.

11. Kopp, R. E., and Oxford, R. J., "Linear Regression Applied to System Identification for Adaptive Control Systems," *AIAA Journal*, Vol. 1, No. 10, pp. 2300–2306, October 1963.

12. Cox, H., "Estimation of State Variables via Dynamic Programming," Preprints of 1964 Joint Automatic Control Conference at Stanford University, pp. 376–381, June 1964.

13. Gallager, R. G., "A Simple Derivation of the Coding Theorem and Some Applications" (to appear in *IEEE Transactions on Information Theory*).

14. Check, M. H., and Reiffen, B., "A Note on Sequential Decoding Applied to Large Alphabet Incoherent Channels—R_{comp} Determination," *MIT Lincoln Lab Report*, Vol. 65, G–6, May 31, 1963.

15. Lebow, I. L., "A Qualitative Description of Sequential Decoding," *MIT Lincoln Lab Report*, Vol. 62, G–4, July 12, 1963.

16. Wozencraft, J. M., and Kennedy, R. S., "Modulation and Demodulation for Probabilistic Coding," (to appear in *IEEE Transactions on Information Theory*).

References (Cont'd)

17. Wozencraft, J. M., and Jacobs, I. M., *Principles of Communication Engineering* (to be published by John Wiley & Sons, New York), 1965.

18. Blustein, G., and Jordan, K. L., Jr., "An Investigation of the Fano Sequential Decoding Algorithm by Computer Simulation," *MIT Lincoln Lab Report*, Vol. 62, G-5, July 12, 1963.